# (19) World Intellectual Property Organization International Bureau





#### (43) International Publication Date 4 July 2002 (04.07,2002)

# PCT

# (10) International Publication Number WO 02/052015 A2

- (51) International Patent Classification<sup>7</sup>: C12N 15/51, 15/40, C12Q 1/68, 1/70, C12N 5/10, 7/04, 15/85
- (21) International Application Number: PCT/CA01/01843
- (22) International Filing Date:

20 December 2001 (20.12.2001)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data: 60/257,857

22 December 2000 (22.12.2000) US

- (71) Applicant (for all designated States except US): BOEHRINGER INGELHEIM (CANADA) LTD. [CA/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): KUKOLJ, George [CA/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA). PAUSE, Arnim [DE/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).

- (74) Agent: BERNIER, Louise, G.; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

#### Published:

 without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: SELF-REPLICATING RNA MOLECULE FROM HEPATITIS C VIRUS

	CLONE APGK-12	AMINO ACID SUBSTITUTIONS								
	5' HCV		EMCV	T		V NS2→6B				3'HCV
	IRES	NeoR-	IRES	NS2	NS3	4A	NS4B	NS5A	NS5B	UTR
77 cfu/µg	G (nt1) SEQ ID NO 1		,							
	A (nt1) SEQ ID NO 24			•	-	-	-	-	-	
86 ch/µg	R3 rep A(nt1) SEQ ID NO 25				R(1135)K S(1560)G	K(1691)R	•	T(1993)A G(2042)C L(2155)P P(2166)L		
2000000cfu/µg	G(nt1) SEQ ID NO 7				R(1135)K S(1560)G	K(1691)R		T(1993)A G(2042)C L(2155)P P(2166)L		

(57) Abstract: A unique HCV RNA molecule is provided having an enhanced efficiency of establishing cell culture replication. Novel adaptive mutations have been identified within the HCV non-structural region that improves the efficiency of establishing persistently replicating HCV RNA in cell culture. This self-replicating polynucleotide molecule contains, contrary to all previous reports, a 5'-NTR that can be either an A as an alternative to the G already disclosed and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule. The G-->A mutation gives rise to HCV RNA molecules that, in conjunction with mutations in the HCV non-structural region, such as the G(2042)C/R mutations, possess greater efficiency of transduction and/or replication. These RNA molecules when transfected in a cell line are useful for evaluating potential inhibitors of HCV replication.



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

1

#### SELF-REPLICATING RNA MOLECULE FROM HEPATITIS C VIRUS

#### FIELD OF THE INVENTION

The present invention relates generally to a HCV RNA molecule that self-replicates in appropriate cell lines, particularly to a self-replicating HCV RNA construct having an enhanced efficiency of establishing cell culture replication.

#### **BACKGROUND OF THE INVENTION**

Hepatitis C virus (HCV) is the major etiological agent of post-transfusion and community-acquired non-A non-B hepatitis worldwide. It is estimated that over 200 million people worldwide are infected by the virus. A high percentage of carriers become chronically infected and many progress to chronic liver disease, so called chronic hepatitis C. This group is in turn at high risk for serious liver disease such as liver cirrhosis, hepatocellular carcinoma and terminal liver disease leading to death.

The mechanism by which HCV establishes viral persistence and causes a high rate of chronic liver disease has not been thoroughly elucidated. It is not known how HCV interacts with and evades the host immune system. In addition, the roles of cellular and humoral immune responses in protection against HCV infection and disease have yet to be established.

20

25

Various clinical studies have been conducted with the goal of identifying pharmaceutical compounds capable of effectively treating HCV infection in patients afflicted with chronic hepatitis C. These studies have involved the use of interferonalpha, alone and in combination with other antiviral agents such as ribavirin. Such studies have shown that a substantial number of the participants do not respond to these therapies, and of those that do respond favorably, a large proportion were found to relapse after termination of treatment. To date there are no broadly effective antiviral compounds for treatment of HCV infection.

30 HCV is an enveloped positive strand RNA virus in the *Flaviviridae* family. The single strand HCV RNA genome is of positive polarity and comprises one open reading frame (ORF) of approximately 9600 nucleotides in length, which encodes a linear polyprotein of approx. 3010 amino acids. In infected cells, this polyprotein is cleaved at multiple sites by cellular and viral proteases to produce structural and non-structural (NS) proteins. The structural proteins (C, E1, E2 and E2-p7) comprise

10

15

20

25

30

35

polypeptides that constitute the virus particle (Hijikata et al., 1991; Grakoui et al., 1993(a) ). The non-structural proteins (NS2, NS3, NS4A, NS4B, NS5A, NS5B) encode for enzymes or accessory factors that catalyze and regulate the replication of the HCV RNA genome. Processing of the structural proteins is catalyzed by host cell proteases (Hijikata et al., 1991). The generation of the mature non-structural proteins is catalyzed by two virally encoded proteases. The first is the NS2/3 zincdependent metalloprotease which auto-catalyses the release of the NS3 protein from the polyprotein. The released NS3 contains a N-terminal serine protease domain (Grakoui et al., 1993(b); Hijikata et al., 1993) and catalyzes the remaining cleavages from the polyprotein. The released NS4A protein has at least two roles. First, forming a stable complex with NS3 protein and assisting in the membrane localization of the NS3/NS4A complex (Kim et al., Arch Virol. 1999, 144: 329-343) and second, acting as a cofactor for NS3 protease activity. This membraneassociated complex, in turn catalyzes the cleavage of the remaining sites on the polyprotein, thus effecting the release of NS4B, NS5A and NS5B (Bartenschlager et al., 1993; Grakoui et al., 1993(a); Hijikata et al., 1993; Love et al., 1996; reviewed in Kwong et al., 1998). The C-terminal segment of the NS3 protein also harbors nucleoside triphosphatase and RNA helicase activity (Kim et al., 1995). The function of the protein NS4B is unknown. NS5A, a highly phosphorylated protein, seems to be responsible for the Interferon resistance of various HCV genotypes (Gale Jr. et al. 1997 Virology 230, 217; Reed et al., 1997. NS5B is an RNA-dependent RNA polymerase (RdRp) that is involved in the replication of HCV.

The open reading frame of the HCV RNA genome is flanked on its 5' end by a non-translated region (NTR) of approx. 340 nucleotides that functions as the internal ribosome entry site (IRES), and on its 3' end by a NTR of approximately 230 nucleotides. Both the 5' and 3' NTRs are important for RNA genome replication. The genomic sequence variance is not evenly distributed over the genome and the 5'NTR and parts of the 3'NTR are the most highly conserved portions. The authentic, highly conserved 3'NTR is the object of US patent 5,874,565 granted to Rice *et al.* 

The cloned and characterized partial and complete sequences of the HCV genome have also been analyzed with regard to appropriate targets for a prospective antiviral therapy. Four viral enzyme activities provide possible targets such as (1) the NS2/3 protease; (2) the NS3/4A protease complex, (3) the NS3 Helicase and (4) the NS5B

20

25

30

35

RNA-dependent RNA polymerase. The NS3/4A protease complex and the NS3 helicase have already been crystallized and their three-dimensional structure determined (Kim *et al.*, 1996; Yem *et al.*, 1998; Love *et al.*, 1996; Kim *et al.*, 1998; Yao *et al.*, 1997; Cho *et al.*, 1998). The NS5B RNA dependent RNA polymerase has also been crystallized to reveal a structure reminiscent of other nucleic acid polymerases (Bressanelli *et al.* 1999, Proc. Natl. Acad. Sci, USA 96: 13034-13039; Ago *et al.* 1999, Structure 7: 1417-1426; Lesburg *et al.* 1999, Nat. Struct. Biol. 6: 937-943).

Even though important targets for the development of a therapy for chronic HCV infection have been defined with these enzymes and even though a worldwide intensive search for suitable inhibitors is ongoing with the aid of rational drug design and HTS, the development of therapy has one major deficiency, namely the lack of cell culture systems or simple animal models, which allow direct and reliable propagation of HCV viruses. The lack of an efficient cell culture system is still the main reason to date that an understanding of HCV replication remains elusive.

Although flavi- and pestivirus self-replicating RNAs have been described and used for the replication in different cell lines with a relatively high yield, similar experiments with HCV have not been successful to date (Khromykh *et al.*, 1997; Behrens *et al.*, 1998; Moser *et al.*, 1998). It is known from different publications that cell lines or primary cell cultures can be infected with high-titer patient serum containing HCV (Lanford *et al.* 1994; Shimizu *et al.* 1993; Mizutani *et al.* 1996; Ikda *et al.* 1998; Fourner *et al.* 1998; Ito *et al.* 1996). However, these virus-infected cell lines or cell cultures do not allow the direct detection of HCV-RNA or HCV antigens.

It is also known from the publications of Yoo *et al.* **1995**; and of Dash *et al.*, **1997**; that hepatoma cell lines can be transfected with synthetic HCV-RNA obtained through *in vitro* transcription of the cloned HCV genome. In both publications the authors started from the basic idea that the viral HCV genome is a plus-strand RNA functioning directly as mRNA after being transfected into the cell, permitting the synthesis of viral proteins in the course of the translation process, and so new HCV particles could form HCV viruses and their RNA detected through RT-PCR. However the published results of the RT-PCR experiments indicate that the HCV replication in the described HCV transfected hepatoma cells is not particularly

4

efficient and not sufficient to measure the quality of replication, let alone measure the modulations in replication after exposure to potential antiviral drugs. Furthermore it is now known that the highly conserved 3' NTR is essential for the virus replication (Yanagi *et al.*, **1999**). This knowledge strictly contradicts the statements of Yoo *et al.* (*supra*) and Dash *et al.* (*supra*), who used for their experiments only HCV genomes with shorter 3' NTRs and not the authentic 3' end of the HCV genome.

In WO 98/39031, Rice *et al.* disclosed authentic HCV genome RNA sequences, in particular containing: a) the highly conserved 5'-terminal sequence "GCCAGCC"; b) the HCV polyprotein coding region; and c) 3'-NTR authentic sequences.

In WO 99/04008, Purcell *et al.* disclosed an HCV infectious clone that also contained only the highly conserved 5'-terminal sequence "GCCAGC".

Recently Lohman *et al.* **1999** (Science 285: 110-113) and Bartenschlager *et al.* (in CA 2,303,526, laid-open on October 3, 2000) disclosed a HCV cell culture system where the viral RNA (I377/NS2-3') self-replicates in the transfected cells with such efficiency that the quality of replication can be measured with accuracy and reproducibility. The Lohman and Bartenschlager disclosures were the first demonstration of HCV RNA replication in cell culture that was substantiated through direct measurement by Northern blots. This replicon system and sequences disclosed therein highlight once again the conserved 5' sequence "GCCAGC". A similar observation highlighting the conservation of the 5'NTR was made by Blight *et al.* **2000** (Science 290: 1972-1974) and WO 01/89364 published on Nov. 29, 2001.

25

30

35

10

In addition to the conservation of the 5' and 3' untranslated regions in cell culture replicating RNAs, three other publications by Lohman *et al.* **2001**, Krieger *et al.* **2001** and Guo *et al.* **2001** have recently disclosed distinct adaptive mutants within the HCV non-structural protein coding region. Specific nucleotide changes that after the amino acids of the HCV non-structural proteins are shown to enhance the efficiency of establishing stable replicating HCV subgenomic replicons in culture cells.

Applicant has now found that, contrary to all previous reports, the highly conserved 5'-NTR can be mutated by adaptation to give rise to a HCV RNA sequence that, in conjunction with mutations in the HCV non-structural region, provides for a greater

5

efficiency of transduction and/or replication.

Applicant has also identified novel adaptive mutations within the HCV non-structural region that improves the efficiency of establishing persistently replicating HCV RNA in cell culture.

One advantage of the present invention is to provide an alternative to these existing systems comprising a HCV RNA molecule that self-replicates. Moreover, the present invention demonstrates that the initiating nucleotide of the plus-strand genome can be either an A as an alternative to the G already disclosed.

A further advantage of the present invention is to provide a unique HCV RNA molecule that transduces and/or replicates with higher efficiency. The Applicant demonstrates the utility of this specific RNA molecule in a cell line and its use in evaluating a specific inhibitor of HCV replication.

#### SUMMARY OF THE INVENTION

In a first embodiment, the present invention provides a 5'-non translated region of
the hepatitis C virus wherein its highly conserved guanine at position 1 is substituted
for adenine.

Particularly, the present invention provides a hepatitis C virus polynucleotide comprising adenine at position 1 as numbered according to the I377/NS2-3' construct (Lohmann et al. 1999, Accession # AJ242651).

Particularly, the invention provides a HCV self-replicating polynucleotide comprising a 5'-terminus consisting of ACCAGC (SEQ ID NO. 8).

In a second embodiment, the present invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.

25

10

6

Particularly, the invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as described above, further comprising the amino acid substitution E(1202)G.

More particularly, the invention provides a HCV self-replicating polynucleotide encoding a polyprotein comprising a G2042C or a G2042R mutation.

Most particularly, the invention provides for HCV self-replicating polynucleotide comprising a nucleotide substitution G-->A at position 1, and said polynucleotide encodes a polyprotein further comprising a G2042C or a G2042R mutation.

Particularly, the polynucleotide of the present invention can be in the form of RNA or DNA that can be transcribed to RNA.

15 In a third embodiment, the invention also provides for an expression vector comprising a DNA form of the above polynucleotide, operably linked with a promoter.

According to a fourth embodiment, there is provided a host cell transfected with the self-replicating polynucleotide or the vector as described above.

20

10

In a fifth embodiment, the present invention provides a RNA replication assay comprising the steps of:

- incubating the host cell as described above in the absence or presence of a potential hepatitis C virus inhibitor;
- isolating the total cellular RNA from the cells;
  - analyzing the RNA so as to measure the amount of HCV RNA replicated;
  - comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor.
- 30 In a sixth embodiment, the invention is directed to a method for testing a compound for inhibiting HCV replication, including the steps of:
  - a) treating the above described host cell with the compound;
  - b) evaluating the treated host cell for reduced replication, wherein reduced replication indicates the ability of the compound to inhibit replication.

7

#### **DETAILED DESCRIPTION OF THE DRAWINGS**

Figure 1 is a schematic view of the bi-cistronic replicon RNA. The sequence deviations between the I377/NS2-3' replicon from Lohman et al., 1999 and the
APGK12 replicon are indicated below the replicon. In place of a G nucleotide at the +1 position in the I377/NS2-3'replicon, the APGK12 contains an additional G resulting in GG at the 5' terminus (the first G being counted as position –1). In the linker region between the neo gene and the EMCV IRES sequence two areas deviate from I377/NS2-3': 14 nucleotides (CGCGCCCAGATGTT) which are not present in I377/NS2/3 are inserted at position 1184 in APGK12; 11 nucleotides (1231-1241) present in I377/NS2-3' are deleted to generate APGK-12. In the NS5B coding region, a T at position 8032 was mutated to C to eliminate a Ncol restriction site.

15 Figure 2 shows Northern blots of RNA-transfected Huh-7 cell lines. 12 µg of total cellular RNA or control RNA was separated on 0.5% agarose-formaldehyde gels and transferred to Hybond N+ paper, fixed and (Figure 2A) radioactively probed with HCV specific minus-strand RNA that detects the presence of plus-strand replicon RNA. Lanes 1 and 2: positive controls that contain 109 copies of in vitro transcribed APGK12 RNA, Lane 3: negative control of total cellular RNA from untransfected 20 Huh-7 cells. Lanes 4 and 5: cellular RNA from B1 and B3 cell lines that have integrated DNA copies of the neomycin phosphotransferase gene. Lane 6: total cellular RNA from a Huh-7 cell line, designated S22.3, that harbors high copy number HCV sub-genomic replicon RNA as highlighted by the arrow. Other cell lines 25 have no detectable replicon RNA. Figure 2B is identical to Figure 2A with the exception that the blot was radioactively probed with HCV specific plus-strand RNA to detect the presence of HCV minus-strand RNA. Lanes 1 and 2 are positive control lanes that contain 109 copies of full length HCV minus strand RNA. Lane 6, which contains 12 µg of total cellular RNA from cell line S22.3, harbors detectable minus-30 strand replicon RNA at the expected size of 8 – 9 kilobases. M represent the migration of non-radioactive molecular size markers on the agarose gel. 28s represents the migration of 28s ribosomal RNA and accounts for the detection of this species in a samples of total cellular RNA.

Figure 3 shows indirect immunofluorescence of a HCV non-structural protein in the

35

8

S22.3 cell line. Indirect immunofluorescence was performed on cells that were cultured and fixed, permeabilized and exposed to a rabbit polyclonal antibody specific for a segment of the HCV NS4A protein. Secondary goat anti-rabbit antibody conjugated with red-fluor Alexa 594 (Molecular Probes) was used for detection. Top panels shows the results of immunofluorescence (40X objective) and the specific staining of the S22.3 cells. The bottom panels represent the identical field of cells viewed by diffractive interference contrast (DIC) microscopy. The majority of S22.3 (Figure 3A) cells within the field stain positively for HCV NS4A protein that localizes in the cytoplasm, whereas the B1 cells (Figure 3B) that fail to express any HCV proteins, only have background level of staining.

10

15

20

Figure 4 shows Western-blots following SDS-PAGE separation of total proteins extracted from three cell lines: (i) naïve Huh-7 cell line, (ii) neomycin resistant Huh-7 cell line B1, and (iii) the S22.3 cell line. Panels A, B, and C, demonstrate the results of western blots probed with rabbit polyclonal antisera specific for neomycin phosphotransferase (NPT), HCV NS3, and HCV NS5B, respectively. Visualization was achieved through autoradiographic detection of a chemiluminescent reactive secondary \ goat anti-rabbit antibody. Panel A shows that the S22.3 RNA replicon cell line, expresses the NPT protein at levels higher than control B1 cells and that the naïve Huh-7 cell line does not produce the NPT protein. Panels B and C show that only the S22.3 cell line produces the mature HCV NS3 and NS5B proteins, respectively. M represents molecular weight (in kilodaltons) of pre-stained polypeptide markers.

Figure 5A and 5B identify the nucleotide and amino acid sequences respectively that differ from the APGK12 sequence in the different HCV bi-cistronic replicons. The S22.3 adapted replicon is a first generation replicon selected following the transfection of RNA transcribed from the APGK12 template. R3, R7, R16 are second generation replicons that were selected following the transfection of RNA isolated from the S22.3 first generation replicon cell line. Figure 5A: Nucleotide mutations that were characterized in each of the adapted replicons are indicated adjacent to the respective segment of the replicon (IRES, NS3, NS4A, NS5A, and NS5B). Figure 5B: Amino acid numbers are numbered according to the full length HCV poly-protein with the first amino acid in the second cistron corresponding to amino acid 810 in NS2 of I377/NS2-3' construct.

9

Figure 6 depicts the colony formation efficiency of four in vitro transcribed HCV subgenomic bi-cistronic replicon RNAs. The APGK12 serves as the reference sequence; highlighted are the initiating nucleotides of the HCV IRES in each of the constructs and the amino acid differences (from the APGK12 reference sequence) in the HCV non-structural region for the two R3-rep. Note that the in vitro transcribed APGK-12 RNAs that harbor either a 5'G or 5'A form colonies with the same efficiency (ca. 80 cfu/µg in panels A and B) following selection with 0.25 mg/ml G418. RNA isolated from the second generation R3 cell line was reverse transcribed into DNA and cloned into the pAPGK12 vector backbone to generate the R3-rep, which was sequenced and found to encode additional changes that included the L(2155)P substitution in the NS5A segment of the HCV polyprotein (compare R3-rep sequence with the R3 sequence in tables 2 and 3). Various quantities of in vitro transcribed R3-rep-5'A RNA, were transfected into naïve Huh-7 cells to determine a colony formation efficiency of 1.2 X 10<sup>6</sup> cfu/µg of RNA (panel C). Various quantities of R3-rep-5'G were also transfected resulting in a colony formation efficiency of 2 X 10<sup>6</sup> cfu/µg of RNA (panel D).

Figure 7 displays a typical RT-PCR amplification plot (left panel) and the graphical representation of Ct values versus known HCV RNA quantity in a standard curve (right panel). Each of the plotted curves in the left panel, graph the increment of fluorescence reporter signal (delta-Rn) versus PCR cycle number for a predetermined quantity of HCV replicon RNA. The Ct value is obtained by determining the point at which the fluorescence exceeds an arbitrary value (horizontal line). The right panel demonstrates the linear relationship between starting RNA copy number of the predetermined standards (large black dots) and the Ct value. Smaller dots are the Ct values of RNA samples (containing unknown quantity of HCV replicon RNA) from S22.3 cells treated with various concentrations of a specific inhibitor of HCV replication.

30

35

20

25

10

**Figure 8** shows the effect of increasing concentration of inhibitor A on HCV RNA replicon levels in Huh7 cells. S22.3 cells were grown in the presence of increasing concentrations of inhibitor A starting at 0.5nM and ranging to 1024nM. The inhibitor dose-response curve is the result of 11 concentrations from serial two-fold dilutions (1:1). One control well, without any inhibitor, was also included during the course of

10

the experiment. The cells were incubated for 4 days in a 5% CO<sub>2</sub> incubator at 37 °C. Total cellular RNA was extracted, quantified by optical density. HCV replicon RNA was evaluated by real time RT-PCR and plotted as genome equivalents/µg total RNA as a function of inhibitor concentration

5

10

15

35

#### **Definitions**

Unless defined otherwise, the scientific and technological terms and nomenclature used herein have the same meaning as commonly understood by a person of ordinary skill to which this invention pertains. Generally, the procedures for cell culture, infection, molecular biology methods and the like are common methods used in the art. Such standard techniques can be found in reference manuals such as for example Sambrook *et al.* (1989) and Ausubel *et al.* (1994).

Nucleotide sequences are presented herein by single strand, in the 5' to 3' direction,

from left to right, using the one letter nucleotide symbols as commonly used in the art and in accordance with the recommendations of the IUPAC-IUB Biochemical Nomenclature Commission (1972).

The present description refers to a number of routinely used recombinant DNA (rDNA) technology terms. Nevertheless, definitions of selected examples of such rDNA terms are provided for clarity and consistency.

The term "DNA segment or molecule or sequence", is used herein, to refer to molecules comprised of the deoxyribonucleotides adenine (A), guanine (G), thymine (T) and/or cytosine (C). These segments, molecules or sequences can be found in nature or synthetically derived. When read in accordance with the genetic code, these sequences can encode a linear stretch or sequence of amino acids which can be referred to as a polypeptide, protein, protein fragment and the like.

As used herein, the term "gene" is well known in the art and relates to a nucleic acid sequence defining a single protein or polypeptide. The polypeptide can be encoded by a full-length sequence or any portion of the coding sequence, so long as the functional activity of the protein is retained.

A "structural gene" defines a DNA sequence which is transcribed into RNA and translated into a protein having a specific structural function that constitute the viral particles. "Structural proteins" defines the HCV proteins incorporated into the virus particles namely, core "C", E1, E2, and E2-p7.

"Non-structural proteins", defines the HCV proteins that are not comprised in viral particles namely, NS2, NS3, NS4A, NS5A and NS5B.

10

25

30

35

"Restriction endonuclease or restriction enzyme" is an enzyme that has the capacity to recognize a specific base sequence (usually 4, 5 or 6 base pairs in length) in a DNA molecule, and to cleave the DNA molecule at every place where this sequence appears. An example of such an enzyme is *EcoRI*, which recognizes the base sequence GJAATTC and cleaves a DNA molecule at this recognition site.

"Restriction fragments" are DNA molecules produced by the digestion of DNA with a restriction endonuclease. Any given genome or DNA segment can be digested by a particular restriction endonuclease into at least two discrete molecules of restriction fragments.

"Agarose gel electrophoresis" is an analytical method for fractionating polynucleotide molecules based on their size. The method is based on the fact that nucleic acid molecules migrate through a gel as through a sieve, whereby the smallest molecule has the greatest mobility and travels the farthest through the gel. The sieving characteristics of the gel retards the largest molecules such that, these have the least mobility. The fractionated polynucleotides can be visualized by staining the gel using methods well known in the art, nucleic acid hybridization or by tagging the fractionated molecules with a detectable label. All these methods are well known in the art, specific methods can be found in Ausubel *et al.* (*supra*).

"Oligonucleotide or oligomer" is a molecule comprised of two or more

20 deoxyribonucleotides or ribonucleotides, preferably more than three. The exact size of the molecule will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. An oligonucleotide can be derived synthetically, by cloning or by amplification.

"Sequence amplification" is a method for generating large amounts of a target sequence. In general, one or more amplification primers are annealed to a nucleic acid sequence. Using appropriate enzymes, sequences found adjacent to, or in between the primers are amplified. An amplification method used herein is the polymerase chain reaction (PCR) and can be used in conjunction with the reverse-transcriptase (RT) to produce amplified DNA copies of specific RNA sequences.

"Amplification primer" refers to an oligonucleotide, capable of annealing to a RNA or DNA region adjacent to a target sequence and serving as the initiation primer for DNA synthesis under suitable conditions well known in the art. The synthesized primer extension product is complementary to the target sequence.

The term "domain" or "region" refers to a specific amino acid sequence that defines either a specific function or structure within a protein. As an example herein, is the

12

NS3 protease domain comprised within the HCV non-structural polyprotein. The terms "plasmid" "vector" or "DNA construct" are commonly known in the art and refer to any genetic element, including, but not limited to, plasmid DNA, phage DNA, viral DNA and the like which can incorporate the oligonucleotide sequences, or sequences of the present invention and serve as DNA vehicle into which DNA of the present invention can be cloned. Numerous types of vectors exist and are well known in the art.

The terminology "expression vector" defines a vector as described above but designed to enable the expression of an inserted sequence following transformation or transfection into a host. The cloned gene (inserted sequence) is usually placed under the control of control element sequences such as promoter sequences. Such expression control sequences will vary depending on whether the vector is designed to express the operably linked gene *in vitro* or *in vivo* in a prokaryotic or eukaryotic host or both (shuttle vectors) and can additionally contain transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements, and/or translational initiation and termination sites.

10

.15

20

25

30

35

A host cell or indicator cell has been "transfected" by exogenous or heterologous DNA (e.g. a DNA construct) or RNA, when such nucleic acid has been introduced inside the cell. The transfecting DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transfecting/transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, an example of a stably transfected cell is one in which the transfecting DNA has become integrated into a chromosome and is inherited by daughter cells through chromosome replication. A host cell or indicator cell can be transfected with RNA. A cell can be stably transfected with RNA if the RNA replicates and copies of the RNA segregate to daughter cells upon cell division. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transfecting DNA or RNA. Transfection methods are well known in the art (Sambrook et al., 1989; Ausubel et al., 1994). If the RNA encodes for a genetic marker that imparts an observable phenotype, such as antibiotic resistance, then the stable transfection of replicating RNA can be monitored by the acquisition of such phenotype by the host cell.

As used herein the term "transduction" refers to the transfer of a genetic marker to host cells by the stable transfection of a replicating RNA.

13

The nucleotide sequences and polypeptides useful to practice the invention include without being limited thereto, mutants, homologs, subtypes, quasi-species, alleles, and the like. It is understood that generally, the sequences of the present invention encode a polyprotein. It will be clear to a person skilled in the art that the polyprotein of the present invention and any variant, derivative or fragment thereof, is auto-processed to an active protease.

As used herein, the designation "variant" denotes in the context of this invention a sequence whether a nucleic acid or amino acid, a molecule that retains a biological activity (either functional or structural) that is substantially similar to that of the original sequence. This variant may be from the same or different species and may be a natural variant or be prepared synthetically. Such variants include amino acid sequences having substitutions, deletions, or additions of one or more amino acids, provided the biological activity of the protein is conserved. The same applies to variants of nucleic acid sequences which can have substitutions, deletions, or additions of one or more nucleotides, provided that the biological activity of the sequence is generally maintained.

10

15

20

30

35

The term "derivative" is intended to include any of the above described variants when comprising additional chemical moiety not normally a part of these molecules. These chemical moieties can have varying purposes including, improving a molecule's solubility, absorption, biological half life, decreasing toxicity and eliminating or decreasing undesirable side effects. Furthermore, these moieties can be used for the purpose of labeling, binding, or they may be comprised in fusion product(s). Different moieties capable of mediating the above described effects can be found in *Remington's The Science and Practice of Pharmacy* (1995).

Methodologies for coupling such moieties to a molecule are well known in the art.

The term "fragment" refers to any segment of an identified DNA, RNA or amino acid sequence and/or any segment of any of the variants or derivatives described herein above that substantially retains its biological activity (functional or structural) as required by the present invention.

The terms "variant", "derivative", and "fragment" of the present invention refer herein to proteins or nucleic acid molecules which can be isolated/purified, synthesized chemically or produced through recombinant DNA technology. All these methods are well known in the art. As exemplified herein below, the nucleotide sequences and polypeptides used in the present invention can be modified, for example by *in vitro* mutagenesis.

10

20

35

As used herein, the term "HCV polyprotein coding region" means the portion of a hepatitis C virus that codes for the polyprotein open reading frame (ORF). This ORF may encode proteins that are the same or different than wild-type HCV proteins. The ORF may also encode only some of the functional protein encoded by wild-type polyprotein coding region. The protein encoded therein may also be from different isolates of HCV, and non-HCV protein may also be encoded therein.

As used herein, the abbreviation "NTR" used in the context of a polynucleotide molecule means a non-translated region. The term "UTR" means untranslated region. Both are used interchangeably.

#### **Preferred embodiments**

Particularly, the invention provides a HCV self-replicating polynucleotide molecule comprising a 5'-terminus consisting of ACCAGC (SEQ ID NO.8).

According to the first embodiment of this invention, there is particularly provided a HCV polynucleotide construct comprising:

- a 5'-non translated region (NTR) comprising the sequence ACCAGC at, or proximal to, its 5'-terminus;
  - a HCV polyprotein coding region; and
  - a 3'-NTR region.

In a second embodiment, the present invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.

Particularly, the invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as described above, further comprising the amino acid substitution E(1202)G.

Alternatively, the first embodiment of the present invention is directed to HCV selfreplicating polynucleotide molecule comprising a G2042C/R mutation. 5

According to the second embodiment, the present invention particularly provides a HCV polynucleotide construct comprising:

- a 5'-NTR region comprising the sequence ACCAGC at, or proximal to, its 5'-terminus;
- a HCV polyprotein region coding for a HCV polyprotein comprising a
   G(2042)C or a G(2042)R mutation; and
- a 3'-NTR region.
- Preferably, the polynucleotide construct of the present invention is a DNA or RNA molecule. More preferably, the construct is a RNA molecule. Most preferably, the construct is a DNA molecule.
- More particularly, the first embodiment of this invention is directed to a RNA

  molecule encoded by the DNA molecule selected from the group consisting of: SEQ

  ID NO. 2, 4, 5, 6, 7, 24 and 25.
  - Most particularly, the invention provides a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.

20

In a third embodiment, the invention also is directed to an expression vector comprising DNA forms of the above polynucleotide, operably linked with a promoter.

Preferably, the promoter is selected from the group consisting of: T3, T7 and SP6.

25

According to a fourth embodiment, there is provided a host cell transfected with the self-replicating polynucleotide or vector as described above. Particularly, the host cell is a eukaryotic cell line. More particularly, the eukaryotic cell line is a hepatic cell line. Most particularly, the hepatic cell line is Huh-7.

30

In a fifth embodiment, the present invention provides a RNA replication assay comprising the steps of:

- a) incubating the host cell as described above under conditions suitable for RNA replication;
- b) isolating the total cellular RNA from the cells; and

c) analyzing the RNA so as to measure the amount of HCV RNA replicated.

Preferably, the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the amount of HCV RNA replicated.

Alternatively in this fifth embodiment, the construct comprises a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.

10

15

According to a preferred aspect of the sixth embodiment, the invention is directed to a method for testing a compound for inhibiting HCV replication, including the steps of:

- a) carrying step a) as described in the above assay, in the presence or absence of the compound;
- b) isolating the total cellular RNA from the cells; and
- c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
- d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor,
- wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.

Preferably, the cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.

25

#### **EXAMPLES**

#### **EXAMPLE 1**

Replicon Constructs (APGK-12; Figure 1)

pET9a-EMCV was obtained by ligating an oligonucleotide linker
5' gaattccagatggcgcccagatgttaaccagatccatggcacactctagagtactgtcgac 3' (SEQ ID NO.9) to pET-9a (Novagen) that was cut with EcoRI and Sall to form the vector pET-9a-mod. This linker contains the following restriction sites: EcoRI, AscI, HpaI, NcoI, XbaI, ScaI, Sall. The EMCV IRES was amplified by PCR from the vector pTM1 with primers

17

5' cggaatcgttaacagaccacaacggtttccctc 3' (SEQ ID NO.10) and 5' ggcgtacccatggtattatcgtgtttttca 3' (SEQ ID NO.11) and ligated into pET-9a-mod via EcoRI and NcoI to form pET-9a-EMCV.

- The sequence of HCV NS2 to NS5B followed by the 3'UTR of HCV was obtained from the replicon construct I377/NS2-3' (Lohman *et al.*, 1999; accession number: AJ242651) and synthesized by Operon Technologies Inc. with a T to C change at the Ncol site in NS5B at nucleotide 8032. This sequence was released from an GenOp® vector (Operon Technologies) with Ncol and Scal and transferred into pET-9a-EMCV to form pET-9a-EMCV-NS2-5B-3'UTR.
  - pET-9a-HCV-neo was obtained by amplification of the HCV IRES from a HCV cDNA isolated from patient serum with primers
  - 5' gcatatgaattctaatacgactcactataggccagccccgattg 3' (SEQ ID NO.12) containing a T7 promoter and primer
  - 5' ggcgcccctttggttttctttgaggtttaggattcgtgctcat 3' (SEQ ID NO.13) and amplification of the neomycin phosphotransferase gene from the vector pcDNA 3.1 (Invitrogen) with primers
  - 5' aaagggcgcatgattgaacaagatggattgcacgca 3' (SEQ ID NO.14) and 5' gcatatgttaactcagaagaactcgtcaagaaggcgata 3' (SEQ ID NO.15). These two PCR
  - fragments were mixed and amplified with primers

    5' gcatatgaattctaatacgactcactataggccagccccgattg 3' (SEQ ID NO.16) and
  - 5' gcatatgttaactcagaagaactcgtcaagaaggcgata 3' (SEQ ID NO.15); cut with Eco RI and Hpal and transferred into pET-9a-mod to form pet-9a-HCV-neo. The EMCV-
- NS2-5B-3'UTR was released from pET-9a-EMCV-NS2-5B-3'UTR with HpaI and ScaI and transferred into pet-9a-HCV-neo that was cut with HpaI to form pET-9a-APGK12. This insert was sequenced with specific successive primers using a ABI Prism® BigDye™ Terminator Cycle sequencing kit and analyzed on ABI Prism® 377 DNA Sequencer and is shown in SEQ ID NO 1.

30

35

15

20

## RNA in vitro transcription

pET-9a-APGK12 DNA was cut with Scal for expression of the full-length replicon or with Bglll for expression of a truncated negative control RNA. DNA was analyzed on a 1% agarose gel and purified by Phenol/Chloroform extraction. RNA was produced using a T7 Ribomax® kit (Promega) followed by extraction with phenol/chloroform

and precipitation with 7.5 M LiCl<sub>2</sub>. RNA was treated with DNAse I for 15 min to remove the DNA template and further purified with an RNeasy® column (Qiagen). RNA integrity was verified on a denaturing formaldehyde 1% agarose gel.

#### 5 EXAMPLE 2

10

15

20

25

30

Primary transfection of Huh7 cells and selection of replicon cell lines Human hepatoma Huh7 cells (Health Science Research Resources Bank, Osaka, Japan) were grown in 10% FBS/DMEM. Cells were grown to 70% confluency, trypsinized, washed with phosphate buffered saline (PBS) and adjusted to 1x10<sup>7</sup> cells/ml of PBS. 800 μl of cells were transferred into 0.4cm cuvettes and mixed with 15 μg of replicon RNA. Cells were electroporated using 960μF, 300 volts for ~18 msec and evenly distributed into two 15 cm tissue culture plates and incubated in a tissue culture incubator for 24 hours. The selection of first and second generation replicon cell lines was with 10% FBS/DMEM medium supplemented with 1mg/ml of G418. Cells were selected for 3-5 weeks until colonies were observed that were isolated and expanded.

Following the G418 selection and propagation of Huh-7 cells transfected with APGK12 (SEQ ID NO. 1) RNA, cells that formed a distinct colony were treated with trypsin and serially passed into larger culture flasks to establish cell lines. Approximately 10 X 10<sup>6</sup> cells were harvested from each cell line. The cells were lysed and the total cellular RNA extracted and purified as outlined in Qiagen RNAeasy® preparatory procedures. Figure 2 shows the analysis of 12 µg of total cellular RNA from various cell lines as analyzed on a Northern blot of a denaturing agarose-formaldehyde gel.

Figure 2A is a Northern blot (radioactively probed with HCV specific minus-strand RNA) that detects the presence of plus-strand replicon RNA. Lanes 1 and 2 are positive controls that contain 10<sup>9</sup> copies *of in vitro* transcribed APGK12 RNA. Lane 2 contains the *in vitro* transcribed RNA mixed with 12 µg of total cellular from naïve Huh-7 cells. Lane 3 is a negative control of total cellular RNA from untreated Huh-7 cells. Lanes 4 and 5 contain cellular RNA from the B1 and B3 G418 resistant cell lines that have DNA integrated copies of the neomycin phosphotransferase gene. Lane 6 contains total cellular RNA from a Huh-7 cell line, designated S22.3, that

harbors high copy number of HCV sub-genomic replicon RNA as detected by the positive signal in the 8 kilo-base range. Other cell lines have no detectable replicon RNA. Figure 2B is a Northern blot of a duplicate of the gel presented in 2A with the exception that the blot was radioactively probed with HCV specific plus-strand RNA to detect the presence of HCV minus-strand RNA (lanes 1 and 2 are positive control lanes that contain  $10^9$  copies of full length genomic HCV minus strand RNA); only lane 6, which contains  $12~\mu g$  of total cellular RNA from cell line S22.3, harbors detectable minus-strand replicon RNA at the expected size of 8-9 kilobases. An quantitative estimation of RNA copy number, based on phosphorimager scanning of the Northern blots, is approximately  $6~X10^7$  copies of plus-strand/ $\mu g$  of total RNA, and  $6~x~10^6$  copies of minus strand/ $\mu g$  of total RNA. The presence of the plus-strand and minus-strand intermediate confirms that the HCV sub-genomic RNA is actively replicating in the S22.3 cell line.

#### 15 EXAMPLE 3

10

20

25

30

35

#### S22.3 cell line constitutively expresses HCV non-structural proteins.

HCV non-structural protein expression was examined in the S22.3 cell line. Figure 3 displays the result of indirect immunofluorescence that detects the HCV NS4A protein in the S22.3 cell line and not in the replicon negative B1 cell line (a G418 resistant Huh-7 cell line). Indirect immunofluorescence was performed on cells that were cultured and fixed (with 4% paraformaldehyde) onto Lab-tek chamber slides. Cells were permeabilized with 0.2% Triton X-100 for 10 minutes followed by a 1 hour treatment with 5% milk powder dissolved in phosphate-buffered saline (PBS). A rabbit serum containing polyclonal antibody raised against a peptide spanning the HCV NS4A region was the primary antibody used in detection. Following a 2 hour incubation with the primary antibody, cells were washed with PBS and a secondary goat anti-rabbit antibody conjugated with red-fluor Alexa® 594 (Molecular Probes) was added to cells for 3 hours. Unbound secondary antibody was removed with PBS washes and cells were sealed with a cover slip. Figure 3 (top panels) shows the results of immunofluorescence as detected by a microscope with specific fluorescent filtering; the bottom panels represent the identical field of cells viewed by diffractive interference contrast (DIC) microscopy. The majority of S22.3 (Figure 3A) cells within the field stain positively for HCV NS4A protein that localizes in the cytoplasm, whereas the B1 cells (Figure 3B) that fail to express any HCV proteins, only have

background level of staining. A small proportion of S22.3 cells express high levels of intensely stained HCV NS4A.

Expression of the proteins encoded by the bi-cistronic replicon RNA was also examined on Western-blots following SDS-PAGE separation of total proteins extracted from: (i) naïve Huh-7 cell line, (ii) neomycin resistant Huh-7 cell line B1, and (iii) the S22.3 cell line. Figure 4 panels A, B, and C, demonstrate the results of western blots probed with rabbit polyclonal antisera specific for neomycin phosphotransferase (NPT), HCV NS3, and HCV NS5B, respectively. Visualization was achieved through autoradiographic detection of a chemiluminescent reactive secondary HRP-conjugated goat anti-rabbit antibody. Figure 4 panel A shows that the S22.3 RNA replicon cell line, expresses the NPT protein at levels higher than B1 cells (which contain an integrated DNA copy of the *npt* gene) and that the naïve Huh-7 cell line does not produce the NPT protein. Figure 4 panels B and C show that only the S22.3 cell line produces the mature HCV NS3 and NS5B proteins, respectively. The western blots demonstrate that the S22.3 cell line, which harbors actively replicating HCV sub-genomic replicon RNA, maintains replication of the RNA through the high level expression of the HCV non-structural proteins.

#### 20 EXAMPLE 4

10

15

25

30

## Sequence determination of adapted replicons

Total RNA was extracted from replicon containing Huh7 cells using a RNeasy Kit (Qiagen). Replicon RNA was reverse transcribed and amplified by PCR using a OneStep RT-PCR kit (Qiagen) and HCV specific primers (as selected from the full-length sequence disclosed in WO 00/66623). Ten distinct RT-PCR products, that covered the entire bi-cistronic replicon in a staggered fashion, were amplified using oligonucleotide primers. The PCR fragments were sequenced directly with ABI Prism® BigDye<sup>TM</sup> Terminator Cycle PCR Sequencing and analyzed on ABI Prism® 377 DNA Sequencer. To analyze the sequence of the HCV replicon 3' and 5' ends a RNA ligation/RT-PCR procedure described in Kolykhalov *et al.* 1996 was followed. The nucleotide sequence of S22.3 is presented as SEQ ID NO. 2.

21

#### EXAMPLE 5.

#### Serial Passage of HCV Replicon RNA

The total cellular RNA from the S22.3 cell line was prepared as described above. HCV Replicon RNA copy number was determined by Tagman® RT-PCR analysis and 20 µg of total S22.3 cellular RNA (containing 1 X 109 copies of HCV RNA) was transfected by electroporation into 8 X 10<sup>6</sup> naïve Huh-7 cells. Transfected cells were subsequently cultured in 10 cm tissue culture plates containing DMEM supplemented with 10% fetal calf serum (10% FCS). Media was changed to DMEM 10 (10% FCS) supplemented with 1 mg/ml G418 24 hours after transfection and then changed every three days. Twenty-three visible colonies formed three to four weeks post-transfection and G418 selection. G418 resistant colonies were expanded into second generation cell lines that represent the first cell lines harboring serially passaged HCV Replicon RNA. Three of these cell lines: R3, R7, and R16 were the 15 subject of further analyses. First, the efficiency of transduction by each of the adapted replicons was determined by electroporation of the total cellular RNA (extracted from the R3, R7 and R16) into naïve Huh-7 cells; following electroporation, the transduction efficiency was determined as described above, by counting the visible G418 resistant colonies that arose following 3 to 5 weeks of 20 G418 selection (Table 1). Second, the sequence of the serially passed adapted replicons was determined from the total cellular RNA that was extracted from each of the R3, R7 and R16 replicon cell lines as described in example 4 (SEQ ID NO. 4, 5, 6). Using the pAPGK12 as a reference sequence (SEQ ID NO. 1), the nucleotide changes that were selected in HCV segment of the adapted replicons are presented 25 in Figure 5A. Some of these nucleotide changes are silent and do not change the encoded amino acid whereas others result in an amino acid substitution. Figure 5B summarizes the amino acid changes encoded by the adapted replicons with the amino acid sequence of pAPGK12 as the reference. It is important to note that the reference sequence APGK-12 (SEQ ID NO.1) contains an extra G at the 5'-terminal 30 (5'-GG) that is not maintained in the replicating RNA of the established cell lines. Also noteworthy is that, in addition to G->A at nucleotide 1, there is also an adapted mutation G->C/R at amino acid 2042 (shown as amino acid 1233 in the sequence listing since a.a. 810 of NS2 is numbered as a.a. 1 in SEQ ID) that can be found in all clones analyzed.

22

TABLE 1
Transfection of Huh-7 cells

	RNA	Copies of Replicon	# Colonies	SEQ ID
5				
	5 ng APKG12 replicon in 20μg total Huh-7 RNA	1.2 x 10 <sup>9</sup>	0	
10	15 μg APKG12 replicon RNA	3 x 10 <sup>12</sup>	1 (S22.3)	1
	20μg total:	. 0		
	S22.3 cellular RNA	3 x 10 <sup>9</sup>	23 (3 clones analyzed)	2
15	R3 cellular RNA	1 x 10 <sup>9</sup>	200	4
	R7 cellular RNA	1 x 10 <sup>9</sup>	20	5
	R16 cellular RNA	3 x 10 <sup>8</sup>	100	6
	cloned R3rep RNA	<u>2.3 x 10<sup>8</sup></u>	2000	7

#### 20 EXAMPLE 6

25

30

# Construction of APGK12 with 5' G-> A substitution (APGK12-5'A, SEQ ID NO.24)

The pAPGK12 DNA was modified to change the first nucleotide in the sequence to replace the 5'GG with a 5'A. The change in the pAPGK12 was introduced by replacing an *EcoRl/Agel* portion of the sequence with a PCR-generated *EcoRl/Agel* fragment that includes the mutation. The oligonucleotides used for the amplification were (SEQ ID. NO. 20): 5'-GTG GAC GAA TTC TAA TAC GAC TCA CTA TAA CCA GCC CCC GAT TGG-3' and (SEQ ID. NO. 21): 5'-GGA ACG CCC GTC GTG GCC AGC CAC GAT-3' and generated a 195 bp DNA fragment that was then digested with *EcoRl* and *Agel*. The resulting 178 bp restriction fragment was used to replace the *EcoRl / Agel* fragment in pAPGK12 to generate the pAPGK12-5'A plasmid.

#### EXAMPLE 7

cDNA cloning of the R3-replicon (R3Rep).

The cDNA clone of the R3 replicon was produced by RT-PCR of RNA extracted from the R3 cell line. The following two oligonucleotides were used: (SEQ ID. NO. 22): 5'-GTC GTC TCT GAC ATG GAG AC-3' and (SEQ ID. NO. 23): 5'-GAG TTG

CTC AGT GGA TTG ATG GGC AGC-3'. The ~4400nt PCR fragment, starting within the NS2 coding region and extending to the 5'-end of the NS5B coding region, was cloned into the plasmid pCR3.1 by TA cloning (Invitrogen). The SacII / XhoI portion of this R3 sequence was then used to replace the SacII / XhoI fragment present in the pAPGK12 and the pAPGK12-5'A described above. Consequently, two R3 cDNA sequences were generated: (I) R3-Rep-5'G with an initiating 5'G (SEQ ID NO.7), and R3-Rep-5'A (SEQ ID NO.25) with an initiating 5'A. Sequencing of the R3 rep cDNA identified unique nucleotide changes that differ from the original pAPGK12 sequence (see Figure 5A); some of these changes are silent and do not change the encoded amino acid, whereas others do result in an amino acid change (see Figure 5B). The differences between R3 and the R3-rep reflect the isolation of a unique R3-rep cDNA clone encoding nucleotide changes that were not observed from the sequencing of the total RNA extracted from the R3 cell line.

#### 15 EXAMPLE 8

10

20

25

30

35

#### Efficiency of colony formation with modified constructs

RNA from pAPGK12, pAPGK12-5'A, pR3-Rep and pR3-Rep-5'A was generated by in vitro transcription using the T7 Ribomax® kit (Promega) as described in example 1 above. The reactions containing the pAPGK12-5'A and pR3-Rep-5'A templates were scaled-up 10-fold due to the limitation of commercial RNA polymerase in initiating transcripts with 5'-A. The full length RNAs and control truncated RNA for each clone were introduced into 8 x 106 naïve Huh-7 cells by electroporation as described in example 2. Replicon RNA was supplemented with total cellular Huh-7 carrier RNA to achieve a final 15-20µg quantity. The cells were then cultured in DMEM medium supplemented with 10% fetal calf serum and 0.25 mg/ml G418 in two 150 mm plates. The lower concentration of G418 was sufficient to isolate and select replicon containing cell lines as none of the transfectants with the control truncated RNA produced any resistant colonies. In contrast, in vitro transcribed APGK-12 RNAs that harbor either a 5'G or 5'A form colonies with the same efficiency (ca. 80 cfu/µg in Figure 6 panels A and B) following selection with G418. Various quantities (ranging from 0.1 ng to 1 µg) of the R3-rep-5'A RNA, were transfected into naïve Huh-7 cells to determine a colony formation efficiency of 1.2 X 10<sup>6</sup> cfu/μg of RNA (Figure 6 panel C depicts transfection with 1 μg of RNA). Various quantities (ranging from 0.1 ng to 1 µg) of R3-rep [5'G] were similarly transfected resulting in a colony formation efficiency of 2 X 10<sup>6</sup> cfu/µg of RNA (Figure 6 panel D

depicts colony formation with 1µg of RNA). Note that, shown for the first time, HCV subgenomic replicons replicate as efficiently with a 5' A nucleotide in place of the 5'G. APGK12 with a 5'A or 5'G RNA have similar transduction efficiencies. Similarly, R3-Rep RNAs with either the 5'A or 5'G both display the markedly increased transduction efficiency. Notably, the adaptive mutants within the HCV non-structural segment encoded by the R3-Rep provides for a substantial increase in transduction efficiency as depicted by the dramatic increase in colony forming units per µg of transfected RNA.

#### 10 EXAMPLE 9

## Quantification of HCV Replicon RNA Levels in Cell lines

S22.3 cells, or cell lines harboring other adapted replicons, were seeded in DMEM supplemented with 10% FBS, PenStrep and 1µg/mL Geneticin. At the end of the incubation period the replicon copy number is evaluated by real-time RT-PCR with 15 the ABI Prism 7700 Sequence Detection System. The TAQMAN® EZ RT-PCR kit provides a system for the detection and analysis of HCV RNA (as first demonstrated by Martell et al. 1999 J. Clin. Microbiol. 37: 327-332). Direct detection of the reverse transcription polymerase chain reaction (RT-PCR) product with no downstream processing is accomplished by monitoring the increase in fluorescence of a dye-20 labeled DNA probe (Figure 6). The nucleotide sequence of both primers (adapted from Ruster, B. Zeuzem, S. and Roth, W.K., 1995. Analytical Biochemistry 224:597-600) and probe (adapted from Hohne, M., Roeske, H. and Schreier, E. 1998, Poster Presentation: P297 at the Fifth International Meeting on Hepatitis C Virus and Related Viruses Molecular Virology and Pathogenesis, Venezia-Lido Italy, June 25-25 28, 1998) located in the 5'-region of the HCV genome are the following:

**HCV** Forward primer:

5' ACG CAG AAA GCG TCT AGC CAT GGC GTT AGT 3' (SEQ ID NO.17)

30

HCV Reverse primer:

5' TCC CGG GGC ACT CGC AAG CAC CCT ATC AGG 3' (SEQ ID NO.18)

**HCV Probe:** 

5' FAM-TGG TCT GCG GAA CGG GTG AGT ACA CC-TAMRA 3' (SEQ ID NO.19)

5 FAM: Fluorescence reporter dye.

TAMRA: Quencher dye.

Using The TAQMAN® EZ RT-PCR kit, the following reaction was set up:

Component	Volume per sample	Final	
	(µL)	Concentration	
RNase-Free Water	16	•	
5X Taqman EZ Buffer	10	1X	
Manganese Acetate 25mM	6	3mM	
dATP 10mM	1.5	300µM	
dCTP 10mM	1.5	300µM	
dGTP 10mM	1.5	300µM	
dUTP 20mM	1.5	300µM	
HCV Forward Primer 10μM	1	200nM	
HCV Reverse Primer 10µM	1	200nM	
HCV Probe 5uM	2	200nM	
rTth DNA Polymerase	2	0.1 <b>U/</b> μL	
2.5U/µL			
AmpErase UNG 1U/μL	0.5	0.01U/µL	
Total Mix	45	-	

10

15

To this reaction mix,  $5\mu L$  of total RNA extracted from S22.3 cells diluted at  $10 ng/\mu L$  was added, for a total of 50 ng of RNA per reaction. The replicon copy number was evaluated with a standard curve made from known amounts of replicon copies (supplemented with 50 ng of wild type Huh-7 RNA) and assayed in an identical reaction mix (Figure 7).

Thermal cycler parameters used for the RT-PCR reaction on the ABI Prism 7700 Sequence Detection System were optimized for HCV detection:

Cycle	Temperature (°C)	Time (Minutes)	Repeat	Reaction
Hold	50	2		Initial Step
Hold	60	30		Reverse
				Transcription
Hold	95	5		<b>UNG Deactivation</b>
O1-	95 -	0:15	2	Melt
Cycle	60	1	2	Anneal/Extend
Cycle	90	0:15	40	Melt
	60	1	40	Anneal/Extend

Quantification is based on the threshold cycle, where the amplification plot crosses a defined fluorescence threshold. Comparison of the threshold cycles provides a highly sensitive measure of relative template concentration in different samples. Monitoring during early cycles, when PCR fidelity is at its highest, provides precise data for accurate quantification. The relative template concentration can be converted to RNA copy numbers by employing a standard curve of HCV RNA with known copy number (Figure 7).

#### 10 Example 10

15

20

A specific HCV NS3 protease anti-viral compound inhibits replication of the HCV replicon in S22.3 cell lines.

In order to determine the effect of a specific HCV NS3 protease anti-viral compound on replicon levels in S22.3 cells, the cells were seeded in 24 Well Cell Culture Cluster at 5 X  $10^4$  cells per well in  $500\mu L$  of DMEM complemented with 10% FBS, PenStrep and  $1\mu g/mL$  Geneticin. Cells were incubated until compound addition in a 5% CO $_2$  incubator at 37 °C. The dose-response curve of the inhibitor displayed 11 concentrations resulting from serial two-fold dilutions (1:1). The starting concentration of compound A was 100nM. One control well (without any compound) was also included in the course of the experiment. The 24 well plates were incubated for 4 days in a 5% CO $_2$  incubator at 37 °C. Following a 4 day incubation period, the cells were washed once with PBS and RNA was extracted with the RNeasy® Mini Kit and Qiashredder® from Qiagen. RNA from each well was eluted in 50uL of  $H_2O$ . The RNA was quantified by optical density at 260nm on a Cary 1E UV-Visible Spectrophotometer. 50 ng of RNA from each well was used to quantify the HCV replicon RNA copy number as detailed in Example 6. The level of inhibition (% inhibition) of each well containing inhibitor was calculated with the following

27

equation (CN = HCV Replicon copy number):

$$\% \cdot inhibition = \left(\frac{CN \cdot control - CN \cdot well}{CN \cdot control}\right) * 100$$

The calculated % inhibition values were then used to determine IC<sub>50</sub>, slope factor (n) and maximum inhibition (I<sub>max</sub>) by the non-linear regression routine NLIN procedure of SAS using the following equation:

$$\% \cdot inhibition = \frac{I_{\text{max}} \times [inhibitor]^n}{[inhibitor]^n + IC_{50}^n}$$

10

Compound A was tested in the assay at least 4 times. The  $IC_{50}$  curves were analyzed individually by the SAS nonlinear regression analysis. Figure 8 shows a typical curve and Table 2 shows the individual and average  $IC_{50}$  values of compound A. The average  $IC_{50}$  of compound A in the replication assay was 1.1nM.

15

 $\frac{\text{TABLE 2}}{\text{IC}_{50}} \text{ of compound A in the S22.3 Cell line Replicon Assay.}$ 

Compound	IC <sub>50</sub> (nM)	Average IC <sub>50</sub> (nM)
	1.2	•
Α	1.2	
	1.0	
	0.9	
		1.1 ± 0.2

#### 20 DISCUSSION

25

The reproducible and robust *ex vivo* propagation of hepatitis C virus, to levels required for the accurate testing of potential anti-viral compounds, has not been achieved with any system. As an alternative approach to studying the molecular mechanisms of hepatitis C virus RNA replication, selectable self-replicating bicistronic RNAs were developed (Lohman *et al.*, 1999, Science 285:110-113; Bartenschlager CA 2,303,526). Minimally, these replicons encode for some or all of

28

the non-structural proteins and also carry a selectable marker such as the neomycin phosphotransferase. Though intracellular steady-state levels of these sub-genomic replicon RNAs among the selected clones is moderate to high, the frequency of generating G418-resistant colonies upon transfection of the consensus RNA described by Lohman et al. or Bartenschlager is very low. Less than 100 colonies are generated when 8 million cells are transfected with 1 µg of in vitro transcribed bicistronic replicon RNA. A low efficiency of colony formation was first noted by Lohmann et al (1999 et al, Science 285:110-113). Since then, Lohmann et al. (2001), Blight et al. (2000), and Guo et al. (2001), have isolated sub-genomic RNAs with markedly improved efficiencies in the colony formation assay. Lohmann et al., 1999 originally reported that selection of sub genomic replicons may not involve the selection of adaptive mutants as serially passaged RNA did not demonstrate an improved transfection efficiency. Nevertheless, in an effort to characterize the function and fitness of replicating HCV RNA, we serially passaged the replicon RNA that was isolated from the first selected cell-line. Notably, a significant increase in colony forming efficiency was obtained from this experiment, even though the quantity of replicon RNA was orders of magnitude lower than originally used to transfect the *in vitro* transcribed RNA. Furthermore, a second round serial passage of replicon RNA from this first generation clone into naive Huh-7 cells provided for yet another increase in colony formation efficiency (Table 1).

Our analysis of replicating HCV RNAs identified several adaptive mutations that enhance the efficiency of colony formation by up to 4 orders of magnitude. Adaptive mutations were found in many non-structural proteins, as well as in the 5' non-translated region. The substitution of the 5'-GG doublet for a 5'-A as the inaugurating nucleotide of the HCV 5'-UTR is a variant of the HCV genome that has not been previously described, despite the sequencing of innumerable genotypes and subtypes from across the world. Our original replicon that carried a 5'-GG evolved to variants with either a single 5'-A or 5'-G, both of which showed equal transduction efficiency. We describe here the first report of a HCV genome that can tolerate and stably maintain a 5'A extremity. Moreover, we were successful in re-introducing this defined single nucleotide substitution into our cDNA clone and generate *in vitro* transcribed RNA harboring such an extremity to confirm that a 5'A functions as efficiently as a 5'G.

10

15

20

25

30

We have identified adaptive amino acid substitutions in the HCV non-structural proteins NS3, NS4A and NS5A in the R3 replicon, and a substitution in NS5B in the R7 clone (see Figure 5B). These mutations, particularly the combination defined by the R3-rep (SEQ ID NO. 7), when reconstituted into a cDNA clone and transcribed onto a RNA replicon, result in a significantly enhanced transduction efficiency of up to 20,000 fold from the original wild type APGK12 replicon RNA. However, the steady state levels of intracellular replicon RNA were comparable from each of the different isolated clones. This result suggests that the increase in replication efficiency by the adaptive mutations does not result in higher stable intracellular RNA levels due to higher RNA replication, but rather confers increased permissivity for establishing the replicon in a greater number of Huh7 cells. Such a phenotype may be manifested transiently, through an initial increase of the amount of *de novo* replication, that is required to surpass a defined threshold to establish persistently replicating RNAs within a population of dividing cells.

Recently three other groups also identified other distinct adaptive mutants. Lohmann et al. (2000) reported enhanced transduction efficiencies of up to 10,000 fold with mutations in NS3, NS4B, NS5A and NS5B. Blight et al. (2000) reported an augmentation of transduction efficiencies up to 20,000 fold with a single mutation in NS5A whereas Guo et al. (2001) reported increases in transduction efficiencies of 5,000-10,000 fold with a deletion of a single amino acid in NS5A. The amino acid substitutions that we describe here have not previously been identified as adaptive mutants that enhance the efficiency of RNA transfection and/or replication. One exception is the mutation of E1202G in NS3 that we found in both the R7 and R16 replicons. This adaptation was previously described by Guo et al (2001) and Krieger et al (2001). All other adaptive mutations, without exception, described herein are unpublished.

The development of selectable subgenomic HCV replicons has provided for potential avenues of exploration on HCV RNA replication, persistence, and pathogenesis in cultured cells. However, the low transduction efficiency with the HCV RNA-containing replicons as originally described (Lohmann *et al.*, 1999) showed that it was not a practical system for reverse genetics studies. The adaptive mutants described herein overcome the low transduction efficiency. In light of the recent descriptions of adaptive mutants by other groups, we note that adaptation can be

achieved by distinct mutations in different HCV NS proteins, although the level of adaptation can vary drastically. The replicons encoding adaptive mutants that are described herein are ideally suited for reverse genetic studies to identify novel HCV targets or host cell targets that may modulate HCV RNA replication or HCV replicon RNA colony formation. The adapted and highly efficient replicons are suitable tools for characterizing subtle genotypic or phenotypic changes that affect an easily quantifiable transduction efficiency.

Lastly, we have used our adapted HCV sub genomic replicon cell-line to

demonstrate the proficient inhibition of HCV RNA replication by a specific small molecule inhibitor of the HCV NS3 protease. This is the first demonstration that an antiviral, designed to specifically inhibit one of the HCV non-structural proteins, inhibits HCV RNA replication in cell culture. Moreover, this compound and our S22.3 cell line validate the proposal that RNA replication is directed by the HCV non-structural proteins NS3 to NS5B. The assay that we have described and validated will be extremely useful in characterizing other inhibitors of HCV non-structural protein function in cell culture in a high throughput fashion.

All references found throughout the present disclosure are herein incorporated by reference whether they be found in the following list or not.

# References

20

Ago et al. 1999, Structure 7: 1417-1426

Ausubel et al., 1994, Current Protocols in Molecular Biology, Wiley, New York.

25 Bartenschlager, R. et al., 1993, J. Virol., 67, 3835-3844.

Behrens et al., 1998, J. Virol. 72, 2364

Bressanelli et al. 1999, Proc. Natl. Acad. Sci, USA 96: 13034-13039

Blight et al. 2000, Science 290: 1972-1974

Cho et al., 1998, J. Biol. Chem., 273, 15045

30 Dash et al., 1997, Am. J. Pathol., 151, 363 – 373

Fourner et al. 1998, J. Gen. Virol. 79, 2376

Gale Jr. et al. 1997 Virology 230, 217

Grakoui, A. et al., 1993(a), J. Virol. 67, 1385-1395.

Grakoui A, et al., 1993(b), Proc Natl Acad Sci USA, 90, 10583-7

35 Guo et al. (2001) J. Virol. 8516-8523

31

Hijikata, M. et al., 1991, Proc. Natl. Acad. Sci. USA. 88, 5547-5551.

Hijikata, M. et al., 1993, J. Virol. 67, 4665-4675.

Ikda et al. 1998, Virus Res. 56, 157

Ito et al. 1996, J. Gen. Virol. 77, 1043 - 1054

5 IUPAC-IUB Biochemical Nomenclature Commission, **1972**, Biochemistry, *11*, 1726-1732.

Kolykhalov et al. 1996 J. of Virology, 7, p. 3363-3371

Khromykh et al., 1997, J. Virol. 71, 1497

Kim, D.W. et al., 1995, Biochem. Biophys. Res. Comm., 215, 160-166.

10 Kim et al., 1996, Cell, 87, 343;

Kim et al., 1998, Structure, 6, 89

Kim et al., 1999, Arch. Virol, 144, 329-343.

Krieger et al. (2001) J. Virol. 4614-4624

Kwong AD. et al., 1998, Antiviral Res., 40, 1-18

15 Lanford et al. 1994, Virology 202, 606

Lesburg et al. 1999, Nat. Struct. Biol. 6: 937-943

Lohman et al. 1999, Science 285: 110-113

Lohman et al. (2001) J. Virol. 1437-1449

Love, R.A. et al., 1996, Cell, 87, 331-342

20 Martell et al. 1999 J. Clin. Microbiol. 37: 327-332

Mizutani et al. 1996, J. Virol. 70, 7219 - 7223

Moser et al., 1998, J. Virol. 72, 5318

Reed et al., 1997, J. Virol. 71, 7187

25

Sambrook *et al.*, **1989**, Molecular Cloning – A Laboratory Manual, Cold Spring Harbor Labs.

Shimizu et al. 1993, PNAS, USA, 90, 6037 - 6041

Yanagi et al., 1999, Proc. Natl. Acad. Sci. USA, 96, 2291-95

Yao et al., 1997, Nature Structural Biology, 4, 463

Yem et al., 1998, Protein Science, 7, 837

30 Yoo et al. 1995, J. Virol., 69, 32 – 38

32

#### **CLAIMS**

- 1. A HCV polynucleotide molecule comprising a 5'-non translated region (NTR) wherein guanine at position 1 is substituted for adenine.
- 2. A HCV self-replicating polynucleotide comprising:
  - a 5'-NTR consisting of ACCAGC (SEQ ID NO. 8);
  - a HCV polyprotein region coding for a HCV polyprotein; and
  - a 3'-NTR region.
- 3. The HCV polynucleotide according to claim 2, wherein said polyprotein comprises one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.
- 4. The HCV polynucleotide encoding a polyprotein comprising one or more of the amino acid substitution as defined in claim 3, and further comprising the amino acid substitution E(1202)G.
- 5. The HCV polynucleotide according to claim 3, wherein said substitution is a G2042C or a G2042R mutation.
- 6. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: K(1691)R; and G(2042)C.
- 7. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; and P(2166)L.
- 8. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; L(2155)P; and P(2166)L.
- 9. The HCV polynucleotide according to claim 3, wherein said substitution is selected

from the group consisting of: E(1202)G; I(1984)V; G(2042)C; and M(2992)T.

- 10. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: S(1148)G; E(1202)G; L(1701)F; G(2042)R; and S(2404)P.
- 11. The HCV polynucleotide according to claim 2, wherein said polynucleotide is a RNA molecule encoded by the DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- 12. The HCV polynucleotide according to claim 2, wherein said polynucleotide is a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- **13.** An expression vector comprising a DNA form of the polynucleotide according to claim 2, operably linked to a promoter.
- **14.** A host cell transfected with the self-replicating polynucleotide molecule according to claim 2.
- 15. A host cell according to claim 14, wherein the host cell is a eukaryotic cell line.
- **16.** A host cell according to claim 15, wherein said eukaryotic cell line is a hepatic cell line.
- A host cell according to claim 16, wherein said hepatic cell line is Huh-7.
- 18. A RNA replication assay comprising the steps of:
  - a) incubating the host cell according to claim 14 under conditions suitable for RNA replication;
  - b) isolating the total cellular RNA from the cells; and
  - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
- 19. The assay according to claim 18, wherein the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the amount of HCV RNA replicated.

- 20. The assay according to claim 18, wherein said polynucleotide encodes for a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.
- 21. A method for testing a compound for inhibiting HCV replication, including the steps
  - a) carrying step a) according to claim 18, in the presence or absence of the compound;
  - b) isolating the total cellular RNA from the cells; and
  - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
  - d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor.

wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.

- 22. The method according to claim 21, wherein said cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.
- 23. A HCV polynucleotide molecule comprising:
  - a 5'-NTR region;
  - a HCV polyprotein region coding for a HCV polyprotein comprising one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T; and a 3'-NTR region.
- 24. The HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as defined in claim 24, further comprising the amino acid substitution E(1202)G.
- **25.** The polynucleotide according to claim 24, wherein said substitution is a G2042C or a G2042R mutation.
- 26. The HCV polynucleotide according to claim 24, wherein said substitution is selected

- from the group consisting of: K(1691)R; and G(2042)C.
- 27. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; and P(2166)L.
- 28. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; L(2155)P; and P(2166)L.
- 29. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: E(1202)G; I(1984)V; G(2042)C; and M(2992)T.
- **30.** The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: S(1148)G; E(1202)G; L(1701)F; G(2042)R; and S(2404)P.
- 31. The HCV polynucleotide according to claim 24, wherein said molecule is a RNA molecule encoded by the DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- 32. The HCV polynucleotide according to claim 24, wherein said molecule is a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- **33.** An expression vector comprising a DNA form of the polynucleotide according to claim 24, operably linked to a promoter.
- 34. A host cell transfected with the self-replicating polynucleotide according to claim 24.
- 35. A host cell according to claim 34, wherein the host cell is a eukaryotic cell line.
- **36.** A host cell according to claim 35, wherein said eukaryotic cell line is a hepatic cell line.
- 37. A host cell according to claim 36, wherein said hepatic cell line is Huh-7.

WO 02/052015

- 38. A RNA replication assay comprising the steps of:
  - incubating the host cell according to claim 34 under conditions suitable for RNA replication;
  - isolating the total cellular RNA from the cells; and analyzing the RNA so as to measure the amount of HCV RNA replicated.
- 39. The assay according to claim 38, wherein the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the amount of HCV RNA replicated.
- 40. The assay according to claim 38, wherein said polynucleotide encodes for a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.
- **41.** A method for testing a compound for inhibiting HCV replication, including the steps of
  - a) carrying step a) according to claim 38, in the presence or absence of the compound;
  - b) isolating the total cellular RNA from the cells; and
  - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
  - d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor,

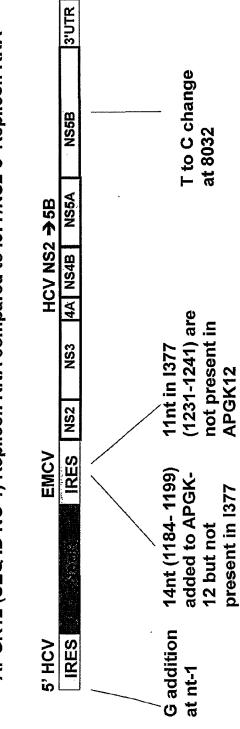
wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.

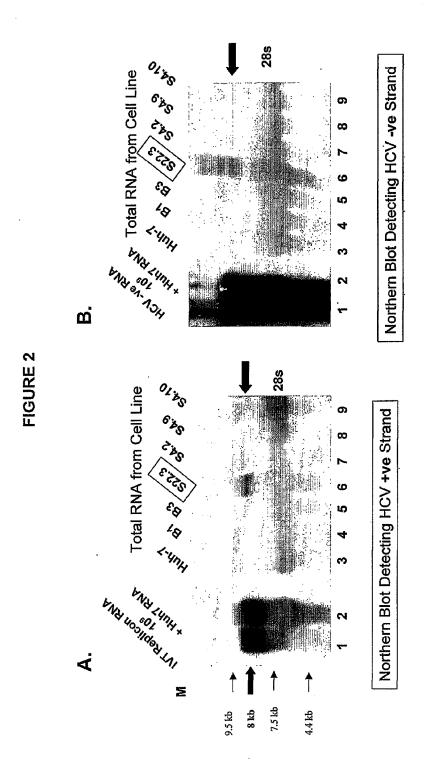
**42.** The method according to claim 41, wherein said cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.

1/9

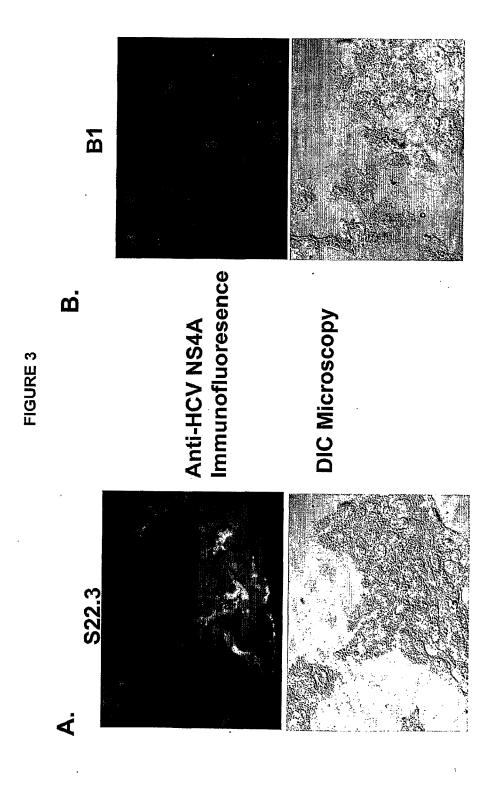
APGK12 (SEQ ID NO 1) Replicon RNA compared to I377/NS2-3' Replicon RNA

FIGURE 1

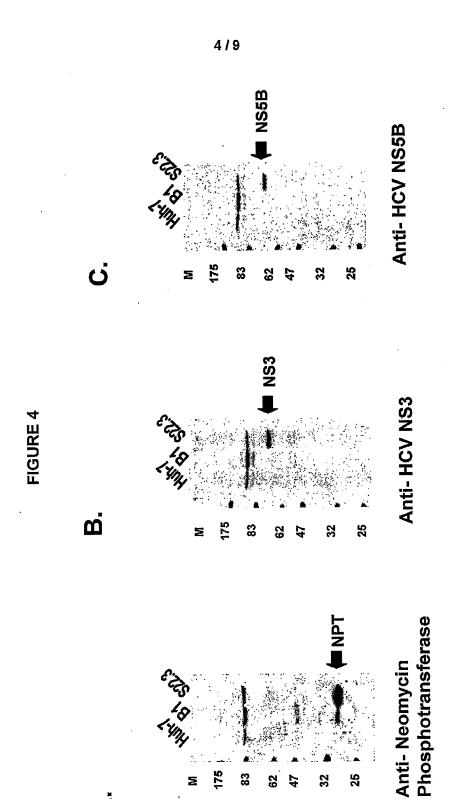




SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)

5/9

**-IGURE 5A** 

	S 22-3 SEQ ID NO 2	R3 SEQ ID NO. 4	R3-rep SEQ ID NO. 7	R7 SEQ ID NO. 5	R16 SEQ ID NO 6
5'end - FIRST nt (HCV IRES)	*G (nt 1) A	G (nt 1) A	•	•	G (nt 1) A
oeN		A (nt 461) G			•
EMCV IRES		A (nt 1739) G	•	•	,
NS 2	•	•			
NS 3	•	G (nt 2778) A A (nt 2840) C A (nt 4052) G	T (nt 2509) C G (nt 2778) A A (nt 2840) C T (nt 3574) C A (nt 4052) G	A (nt 2935) G A (nt 2979) G	A (nt 2816) G A (nt 2979) G
NS 4A	A (nt 4446) R	A (nt 4446) G	C (nt 4387) T A (nt 4446) G C (nt 4507) T	•	C (nt 4475) T
NS 4B	•	T (nt 4855) C	T (nt 4855) C	•	1
NS 5A	G (Nt 5498) T A (Nt 6268) R	A (nt 5351) G G (nt 5488) T G (nt 5659) A C (nt 5871) T A (nt 5268) G	A (nt 5351) G G (nt 5498) T G (nt 5659) A T (nt 5838) C C (nt 5871) T A (nt 6115) G	A (nt 5324) G G (nt 5498) T T (nt 6001) C	G (nt 5498) C T (nt 6320) C T (nt 6584) C
NS 5B	s .	A ( nt 6652) G	•	C (nt 7252) T T (nt 8349) C	•
3'end - last 98 nt	•	9	,		•

rst nt \* G from HCV ires

6/9

G (2042) R S (2404) P S (1148) G R16 SEQ E (1202) G ID NO. 6 L (1701) F G (nt 1) A R3 Rep SEQ ID R7 SEQ ID NO. 5 I (1984) V G (2042) C E (1202) G M (2992) T R (1135) K S (1560) G T (1993) A G (2042) C L (2155) P P (2166) L K (1691) R S 22-3 SEQ ID R3 SEQ ID NO. 2 NO. 4 T (1993) A G (2042) C R (1135) K S (1560) G K (1691) R FIGURE 5B G (nt 1) A P (2166) L K (1691) mix K/R G (2042) C G (nt 1) A 5'end - FIRST nt (HCV IRES) 3'end - last 98 nt NS 5B NS 4A NS 4B NS 5A NS 2 NS 3

first a.a. of NS2 = 810

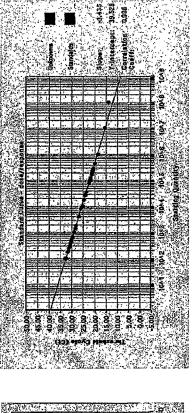
7/9

3"HCV UTR NS5B T(1993)A L(2155)P P(2166)L G(2042)C L(2155)P G(2042)C P(2166)L T(1993)A NS5A AMINO ACID SUBSTITUTIONS NS4B K(1691)R K(1691)R HCV NS2→5B 44 R(1135)K S(1560)G R(1135)K S(1560)G FIGURE 6 NS3 NS2 EMCV IRES NeoR SEQ ID NO 25 SEQ ID NO 24 SEQ ID NO 7 SEQ ID NO 1 CLONE APGK-12 5' HCV IRES G (nt1) R3 rep A (nt1) A(nt1) G(nt1) 1100000cfu/µg 2000000cfu/µg 77 cfu/µg 86 cfu/µg

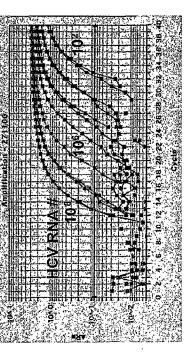
SUBSTITUTE SHEET (RULE 26)

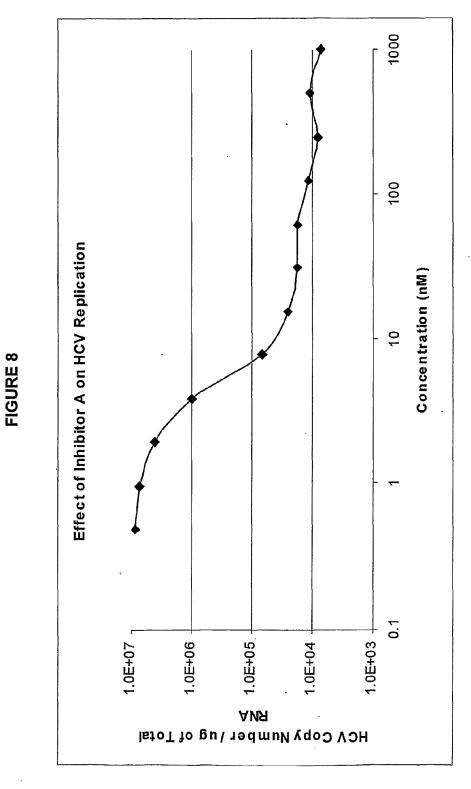
HCV-Replicon: RNA Quantification

FIGURE 7



Ct = Threshold cycle  $\alpha$  Starting RNA Quantity





SUBSTITUTE SHEET (RULE 26)

#### 1/93

#### **SEQUENCE LISTING**

```
<110> BOEHRINGER INGELHEIM (CANADA) LTD.
<120> SELF REPLICATING RNA MOLECULE FROM
 HEPATITIS C VIRUS
<130> 13/083
<150> 60/257,857
<151> 2000-12-22
<160> 25
<170> FastSEQ for Windows Version 4.0
<210> 1
<211> 8639
<212> DNA
<213> HCV
<220>
<221> CDS
<222> (1803)...(8408)
<400> 1
ggccagcccc cgattggggg cgacactcca ccatagatca ctcccctgtg aggaactact 60
gtetteaege agaaagegte tageeatgge gttagtatga gtgtegtgea geeteeagga 120
coccectee egggagagee atagtggtet geggaacegg tgagtacace ggaattgcca 180
ggacgaccgg gtcctttctt ggatcaaccc gctcaatgcc tggagatttg ggcgtgcccc 240
cgcgagactg ctagccgagt agtgttgggt cgcgaaaggc cttgtggtac tgcctgatag 300
ggtgcttgcg agtgccccgg gaggtctcgt agaccgtgca ccatgagcac gaatcctaaa 360
cctcaaagaa aaaccaaagg gcgcgccatg attgaacaag atggattgca cgcaggttct 420
ccqqccqctt qqqtqqaqaq qctattcqqc tatqactqqq cacaacaqac aatcqqctqc 480
tetgatgeeg eegtgtteeg getgteageg eaggggegee eggttetttt tgteaagace 540
qacctqtccq qtqccctgaa tgaactgcag gacgaggcag cgcggctatc gtggctggcc 600
acgacgggcg ttccttgcgc agctgtgctc gacgttgtca ctgaagcggg aagggactgg 660
ctgctattgg gcgaagtgcc ggggcaggat ctcctgtcat ctcaccttgc tcctgccgag 720
aaagtateea teatggetga tgeaatgegg eggetgeata egettgatee ggetaeetge 780
ccattcgacc accaagcgaa acatcgcatc gagcgagcac gtactcggat ggaagccggt 840
cttgtcgatc aggatgatct ggacgaagag catcaggggc tcgcgccagc cgaactgttc 900
gccaggetea aggegegeat gecegaegge gaggateteg tegtgaecea tggegatgee 960
tgcttgccga atatcatggt ggaaaatggc cgcttttctg gattcatcga ctgtggccgg 1020
ctgggtgtgg cggaccgcta tcaggacata gcgttggcta cccgtgatat tgctgaagag 1080
cttggcggcg aatgggctga ccgcttcctc gtgctttacg gtatcgccgc tcccgattcg 1140
cagogcatcg cottotatcg cottottgac gagttottot gagttogcgc ccagatgtta 1200
acagaccaca acggtttccc tctagcggga tcaattccgc cccccccct aacgttactg 1260
gccgaagccg cttggaataa ggccggtgtg cgtttgtcta tatgttattt tccaccatat 1320
tgccgtcttt tggcaatgtg agggcccgga aacctggccc tgtcttcttg acgagcattc 1380
ctaggggtct ttcccctctc gccaaaggaa tgcaaggtct gttgaatgtc gtgaaggaag 1440
cagttcctct ggaagcttct tgaagacaaa caacgtctgt agcgaccctt tgcaggcagc 1500
ggaaccccc acctggcgac aggtgcctct gcggccaaaa gccacgtgta taagatacac 1560
ctgcaaaqqc qqcacaacqc caqtqccacg ttgtgagttg qataqttgtg gaaagagtca 1620
aatggctctc ctcaagcgta ttcaacaagg ggctgaagga tgcccagaag gtaccccatt 1680
gtatgggatc tgatctgggg cctcggtgca catgctttac atgtgtttag tcgaggttaa 1740
aaaacgtcta ggccccccga accacgggga cgtggttttc ctttgaaaaa cacgataata 1800
```

cc atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val 1 5 10 15	1847
ggt etg ata ete ttg ace ttg tea eeg eae tat aag etg tte ete get Gly Leu Ile Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala 20 25 30	
agg ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac Arg Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His 35 40 45	
ttg caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc Leu Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala 50 55 60	
gtc atc ctc ctc acg tgc gcg atc cac cca gag cta atc ttt acc atc Val Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile 65 70 75	
acc aaa atc ttg ctc gcc ata ctc ggt cca ctc atg gtg ctc cag gct Thr Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala 80 85 90	
ggt ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt Gly Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg 100 105 110	
gca tgc atg ctg gtg cgg aag gtt gct ggg ggt cat tat gtc caa atg Ala Cys Met Leu Val Arg Lys Val Ala Gly Gly His Tyr Val Gln Met 115 120 125	
get etc atg aag ttg gee gea etg aca ggt acg tac gtt tat gae eat Ala Leu Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His 130 135 140	
ctc acc cca ctg cgg gac tgg gcc cac gcg ggc cta cga gac ctt gcg Leu Thr Pro Leu Arg Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala 145 150 155	
gtg gca gtt gag ccc gtc gtc ttc tct gat atg gag acc aag gtt atc Val Ala Val Glu Pro Val Val Phe Ser Asp Met Glu Thr Lys Val Ile 160 165 170 175	
ace tgg ggg gca gac ace gcg gcg tgt ggg gac ate ate ttg ggc etg Thr Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Leu Gly Leu 180	
ccc gtc tcc gcc cgc agg ggg agg gag ata cat ctg gga ccg gca gac Pro Val Ser Ala Arg Arg Gly Arg Glu Ile His Leu Gly Pro Ala Asp 195 200 205	
age ett gaa ggg cag ggg tgg ega etc etc geg eet att aeg gee tae Ser Leu Glu Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr 210 215 220	

tcc Ser	caa Gln 225	cag Gln	acg Thr	cga Arg	gly	cta Leu 230	ctt Leu	ggc	tgc Cys	atc Ile	atc Ile 235	act Thr	agc Ser	ctc Leu	aca Thr	2519
	cgg Arg															2567
gca Ala	aca Thr	caa Gln	tct Ser	ttc Phe 260	ctg Leu	gcg Ala	acc Thr	tgc Cys	gtc Val 265	aat Asn	gly ggc	gtg Val	tgt Cys	tgg Trp 270	act Thr	2615
gtc Val	tat Tyr	cat His	ggt Gly 275	gcc Ala	gly ggc	tca Ser	aag Lys	acc Thr 280	ctt Leu	gcc Ala	gly ggc	cca Pro	аад Ъуз 285	gly ggc	cca Pro	2663
atc Ile	acc Thr	caa Gln 290	atg Met	tac Tyr	acc Thr	aat Asn	gtg Val 295	gac Asp	cag Gln	gac Asp	ctc Leu	gtc Val 300	gjy ggc	tgg Trp	caa Gln	2711
	ccc Pro 305															2759
gac Asp 320	ctt Leu	tac Tyr	ttg Leu	gtc Val	acg Thr 325	agg Arg	cat His	gcc Ala	gat Asp	gtc Val 330	att Ile	ccg Pro	gtg Val	cgc Arg	cgg Arg 335	2807
cgg Arg	ggc Gly	gac Asp	agc Ser	agg Arg 340	GJÀ 333	agc Ser	cta Leu	ctc Leu	tcc Ser 345	ccc Pro	agg Arg	ccc Pro	gtc Val	tcc Ser 350	tac Tyr	2855
ttg Leu	aag Lys	ggc Gly	tct Ser 355	tcg Ser	ggc	ggt	cca Pro	ctg Leu 360	ctc Leu	tgc Cys	ccc Pro	tcg Ser	999 Gly 365	cac His	gct Ala	2903
gtg Val	ggc Gly	atc Ile 370	ttt Phe	cgg Arg	gct Ala	gcc Ala	gtg Val 375	tgc Cys	acc Thr	cga Arg	61Å 888	gtt Val 380	gcg Ala	aag Lys	gcg Ala	2951
gtg Val	gac Asp 385	ttt Phe	gta Val	ccc Pro	gtc Val	gag Glu 390	tct Ser	atg Met	gaa Glu	acc Thr	act Thr 395	atg Met	cgg Arg	tcc Ser	ccg Pro	2999
gtc Val 400	ttc Phe	acg Thr	gac Asp	aac Asn	Ser	tcc Ser	cct Pro	ccg Pro	gcc Ala	gta Val 410	ccg Pro	cag Gln	aca Thr	ttc Phe	cag Gln 415	3047
					405					720						
	gcc Ala				gcc					ggc						3095

tcc Ser	gtc Val	gcc Ala 450	gcc Ala	acc Thr	cta Leu	ggt Gly	ttc Phe 455	ej aaa	gcg Ala	tat Tyr	atg Met	tct Ser 460	aag Lys	gca Ala	cat His	3191
ggt Gly	atc Ile 465	gac Asp	cct Pro	aac Asn	atc Ile	aga Arg 470	acc Thr	ej aaa	gta Val	agg Arg	acc Thr 475	atc Ile	acc Thr	acg Thr	ggt Gly	3239
				tac Tyr												3287
tgc Cys	tct Ser	gjà aaa	ggc	gcc Ala 500	tat Tyr	gac Asp	atc Ile	ata Ile	ata Ile 505	tgt Cys	gat Asp	gag Glu	tgc Cys	cac His 510	tca Ser	3335
				act Thr												3383
				gcg Ala												3431
				gtg Val												3479
agc Ser 560	act Thr	gga Gly	gaa Glu	atc Ile	ccc Pro 565	ttt Phe	tat Tyr	ggc	aaa Lys	gcc Ala 570	atc Ile	ccc Pro	atc Ile	gag Glu	acc Thr 575	3527
				agg Arg 580												3575
gat Asp	gag Glu	ctc Leu	gcc Ala 595	gcg Ala	aag Lys	ctg Leu	tcc Ser	Gly Ggc	ctc Leu	gga Gly	ctc Leu	aat Asn	gct Ala 605	gta Val	gca Ala	3623
				ctt Leu												3671
att Ile	gtc Val 625	gta Val	gca Ala	acg Thr	gac Asp	gct Ala 630	cta Leu	atg Met	acg Thr	ggc Gly	ttt Phe 635	acc Thr	ggc Gly	gat Asp	ttc Phe	3719
gac Asp 640	tca Ser	gtg Val	atc Ile	gac Asp	tgc Cys 645	aat Asn	aca Thr	tgt Cys	gtc Val	acc Thr 650	cag Gln	aca Thr	gtc Val	gac Asp	ttc Phe 655	3767
agc Ser	ctg Leu	gac <b>Asp</b>	ccg Pro	acc Thr 660	ttc Phe	acc Thr	att Ile	gag Glu	acg Thr 665	acg Thr	acc Thr	gtg Val	cca Pro	caa Gln 670	gac Asp	3815

gcg Ala	gtg Val	tca Ser	cgc Arg 675	tcg Ser	cag Gln	cgg Arg	cga Arg	ggc 680	agg Arg	act Thr	ggt Gly	agg Arg	ggc Gly 685	agg Arg	atg Met	3863
ggc ggc	att Ile	tac Tyr 690	agg Arg	ttt Phe	gtg Val	act Thr	cca Pro 695	gga Gly	gaa Glu	cgg Arg	ccc Pro	tcg Ser 700	gly	atg Met	ttc Phe	3911
gat Asp	tcc Ser 705	tcg Ser	gtt Val	ctg Leu	tgc Cys	gag Glu 710	tgc Cya	tat Tyr	gac Asp	gcg Ala	ggc Gly 715	tgt Cys	gct Ala	tgg Trp	tac Tyr	3959
gag Glu 720	ctc Leu	acg Thr	ccc Pro	gcc Ala	gag Glu 725	acc Thr	tca Ser	gtt Val	agg Arg	ttg Leu 730	cgg Arg	gct Ala	tac Tyr	cta Leu	aac Asn 735	4007
aca Thr	cca Pro	gly aaa	ttg Leu	ccc Pro 740	gtc Val	tgc Cys	cag Gln	gac Asp	cat His 745	ctg Leu	gag Glu	ttc Phe	tgg Trp	gag Glu 750	agc Ser	4055
gtc Val	ttt Phe	aca Thr	ggc Gly 755	ctc Leu	acc Thr	cac His	ata Ile	gac Asp 760	gcc Ala	cat His	ttc Phe	ttg Leu	tcc Ser 765	cag Gln	act Thr	4103
aag Lys	cag Gln	gca Ala 770	gga Gly	gac Asp	aac Asn	ttc Phe	ccc Pro 775	tac Tyr	ctg Leu	gta Val	gca Ala	tac Tyr 780	cag Gln	gct Ala	acg Thr	4151
gtg Val	tgc Cys 785	gcc Ala	agg Arg	gct Ala	cag Gln	gct Ala 790	cca Pro	cct Pro	cca Pro	tcg Ser	tgg Trp 795	gac Asp	caa Gln	atg Met	tgg Trp	4199
aag Lys 800	tgt Cys	ctc Leu	ata Ile	cgg	cta Leu 805	aag Lys	cct Pro	acg Thr	ctg Leu	cac His 810	G1A aaa	cca Pro	acg Thr	Pro	ctg Leu 815	4247
ctg Leu	tat Tyr	agg Arg	ctg Leu	gga Gly 820	gcc Ala	gtt Val	caa Gln	aac Asn	gag Glu 825	gtt Val	act Thr	acc Thr	aca Thr	cac His 830	ccc Pro	4295
ata Ile	acc Thr	aaa Lys	tac Tyr 835	atc Ile	atg Met	gca Ala	tgc Cys	atg Met 840	tcg Ser	gct Ala	gac Asp	ctg Leu	gag Glu 845	gtc Val	gtc Val	4343
acg Thr	agc Ser	acc Thr 850	tgg Trp	gtg Val	ctg Leu	gta Val	ggc Gly 855	gga Gly	gtc Val	cta Leu	gca Ala	gct Ala 860	ctg Leu	gcc Ala	gcg Ala	4391
tat Tyr	tgc Cys 865	ctg Leu	aca Thr	aca Thr	ggc	agc Ser 870	gtg Val	gtc Val	att Ile	gtg Val	ggc Gly 875	agg Arg	atc Ile	atc Ile	ttg Leu	4439
				gcc		att Ile										4487

						tgc Cys										4535
						caa Gln										4583
						gcg Ala										4631
_					-	gcc Ala 950				_		_				4679
						tta Leu										4727
			-		-	atg Met	_			_				_	_	4775
				Hìs		ctc Leu			Asn					$\operatorname{Trp}$		4823
_	_		Leu	_		ccc Pro		Āla	_		-		Val		_	4871
		Āla			-	gtt Val 1030	Gly	_				Gly	_			4919
	Asp					$\mathtt{Tyr}$					Ala				gtg Val 1055	4967
					Ser	ggc Gly				Ser			Asp		Val	5019
				Āla		ctc Leu			Gly					Gly		5063
			Ala			cgt Arg		His					Glu			5111
		$\mathtt{Trp}$				ctg Leu 1110	Ile					Arg				5159

gtc tcc ccc ac Val Ser Pro Th 1120	g cac tat gtg : His Tyr Val 1125	cct gag agc Pro Glu Ser	gac gct gca g Asp Ala Ala A 1130	ca cgt gtc 5207 la Arg Val 1135
act cag atc ct Thr Gln Ile Le			Gln Leu Leu L	
cac cag tgg at His Gln Trp Il 11	a Asn Glu Asp	tgc tcc acg Cys Ser Thr 1160	Pro Cys Ser G	gc tcg tgg 5303 ly Ser Trp 165
cta aga gat gt Leu Arg Asp Va 1170	tgg gat tgg L Trp Asp Trp	ata tgc acg Ile Cys Thr 1175	gtg ttg act g Val Leu Thr A 1180	at ttc aag 5351 sp Phe Lys
acc tgg ctc ca Thr Trp Leu Gl 1185		Leu Pro Arg		
ttc tca tgt ca Phe Ser Cys Gl 1200	a cgt ggg tac n Arg Gly Tyr 1205	aag gga gtc Lys Gly Val	tgg cgg ggc g Trp Arg Gly A 1210	ac ggc atc 5447 sp Gly Ile 1215
atg caa acc ac Met Gln Thr Th	c tgc cca tgt r Cys Pro Cys 1220	gga gca cag Gly Ala Gln 122	Ile Thr Gly H	at gtg aaa 5495 is Val Lys 1230
aac ggt tcc at Asn Gly Ser Me 12	t Arg Ile Val		Thr Cys Ser A	
cat gga aca tt His Gly Thr Ph 1250				
tec eeg geg ee Ser Pro Ala Pr 1265		Arg Ala Leu		
gag tac gtg ga Glu Tyr Val Gl 1280				
ato acc act da				
Met Thr Thr As				
	p Asn Val Lys 1300 a gtg gat ggg u Val Asp Gly	Cys Pro Cys 130 gtg cgg ttg	Gln Val Pro A 5 cac agg tac g His Arg Tyr A	la Pro Glu 1310 ct cca gcg 5783

Gln		Leu					Leu					Glu	ccg Pro			5879
gca Ala 1360	Val					Leu					His					5927
acg Thr					Leu					Pro					Ser	5975
tca Ser		_	_	Gln	_				Ser	_	_		_	Cys		6023
acc Thr	cgt Arg	cat His 1410	Asp	tcc Ser	ccg Pro	gac Asp	gct Ala 1415	Asp	ctc Leu	atc Ile	gag Glu	gcc Ala 1420	Asn	ctc Leu	ctg Leu	6071
tgg Trp	cgg Arg 1425	Gln	gag Glu	atg Met	ggc	999 Gly 1430	Asn	atc Ile	acc Thr	cgc Arg	gtg Val 1435	Glu	tca Ser	gaa Glu	aat Asn	6119
аад Lys 1440	Val					Ser					${\tt Gln}$					6167
gag Glu					Val					Leu					Lys	6215
. + + -																
Phe				Met					Arg				aac Asn 1485	Pro		6263
	Pro tta	Arg	Ala 1475 tcc Ser	Met 5 tgg	Pro aag	Ile gac	Trp	Ala 1480 gac Asp	Arg ) tac	Pro	Asp	Tyr	Asn 1489 gtg Val	Pro 5 gta	Pro	6263
ctg Leu ggg Gly	Pro tta Leu tgt	gag Glu 1490 cca Pro	Ala 1475 tcc ser	Met tgg Trp	Pro aag Lys cct	Ile gac Asp	Trp ccg Pro 1495 aag Lys	Ala 1480 gac Asp	Arg tac Tyr	Pro gtc Val	Asp cct Pro	cca Pro 1500	Asn 1489 gtg Val ) cct	gta Val	Pro cac His	
ctg Leu ggg Gly	Pro tta Leu tgt Cys 1505 aag Lys	gag Glu 1490 cca Pro	Ala 1475 tcc Ser ) ttg Leu	Met  tgg Trp  ccg Pro	aag Lys cct Pro	gac Asp gcc Ala 1510 ctg Leu	ccg Pro 1495 aag Lys	Ala 1480 gac Asp gcc Ala	tac Tyr cct Pro	gtc Val .ccg Pro	Asp cct Pro ata Ile 1515 gtg Val	cca Pro 1500 cca Pro	Asn 1489 Val Val cct Pro	gta Val cca Pro	Pro cac His cgg Arg	6311
ctg Leu ggg Gly agg	tta Leu tgt Cys 1505 aag Lys	gag Glu 1490 cca Pro agg Arg	tcc ser ttg Leu acg Thr	tgg Trp ccg Pro gtt Val	aag Lys cct Pro gtc Val 1525 aag Lys	gac Asp gcc Ala 1510 ctg Leu	ccg Pro 1495 aag Lys tca Ser	Ala 1480 gac Asp gcc Ala gaa Glu	tac Tyr cct Pro tct Ser	gtc Val .ccg Pro acc Thr 1530	Asp cct Pro ata Ile 1515 gtg Val	cca Pro 1500 cca Pro tct Ser	Asn 1489 ytg Val Cct Pro	gta Val cca Pro gcc Ala	cac His cgg Arg ttg Leu 1535 gtc Val	6311 6359

gac gcg gga tcc Asp Ala Gly Ser 1570				
ggg gag ccg ggg Gly Glu Pro Gly 1585	~ •	Leu Ser Asp		<del>-</del>
agc gag gag gct Ser Glu Glu Ala 1600				
tgg aca ggc gcc Trp Thr Gly Ala	- , -		Ala Glu Glu	<del>-</del> -
ccc atc aat gca Pro Ile Asn Ala 163	Leu Ser Asn	_	Arg His His	
tat gct aca aca Tyr Ala Thr Thr 1650				
tit gac aga ctg Phe Asp Arg Leu 1665		Asp Asp His		
gag atg aag gcg Glu Met Lys Ala 1680				
gag gaa gcc tgt Glu Glu Ala Cys			Ser Ala Arg	
ggc tat ggg gca Gly Tyr Gly Ala 171	Lys Asp Val		Ser Ser Lys	-
cac atc cgc tcc His Ile Arg Ser 1730				
att gac acc acc Ile Asp Thr Thr 1745		Lys Asn Glu		
gag aag ggg ggc Glu Lys Gly Gly 1760				

	atg ggc tct tca tac Met Gly Ser Ser Tyr 1800		
gga cag cgg gtc gag Gly Gln Arg Val Glu 1810	ttc ctg gtg aat gcc Phe Leu Val Asn Ala 1815	tgg aaa gcg aag aaa Trp Lys Ala Lys Lys 1820	tgc 7271 Cys
	tat gac acc cgc tgt Tyr Asp Thr Arg Cys 1830		
	gtt gag gag tca atc Val Glu Glu Ser Ile 1845		
gcc ccc gaa gcc aga Ala Pro Glu Ala Arg 1860	cag gcc ata agg tcg Gln Ala Ile Arg Ser 0 1865	Leu Thr Glu Arg Leu	Tyr
	act aat tot aaa ggg Thr Asn Ser Lys Gly 1880		
	ggt gta ctg acg acc Gly Val Leu Thr Thr 1895		
	gcc gct gcg gcc tgt Ala Ala Ala Ala Cys 1910		
	gta tgc gga gac gac Val Cys Gly Asp Asp 1925		
agc gcg ggg acc caa Ser Ala Gly Thr Gln 1940	gag gac gag gcg agc Glu Asp Glu Ala Ser O 1945	Leu Arg Ala Phe Thr	Glu
	tct gcc ccc cct ggg Ser Ala Pro Pro Gly 1960		
	ata aca tca tgc tcc Ile Thr Ser Cys Ser 1975		
	aaa agg gtg tac tat Lys Arg Val Tyr Tyr 1990		
acc ccc ctt gcg cgg Thr Pro Leu Ala Arg 2000	gct gcg tgg gag aca Ala Ala Trp Glu Thr 2005	gct aga cac act cca Ala Arg His Thr Pro 2010	gtc 7847 Val 2015

## 11/93

								* * /	50							
aat i Asn 8					Asn					Ala					Ala	7895
agg a Arg I				Met					Ser					Gln		7943
caa e Gln l		-	Lys	_		-	-	Gln				-	Cys			7991
att g		Pro					Gln					Leu				8039
agc g Ser 2 2080	Ala.					Ser					Glu					8087
gct (					Lys					Pro					Arg	8135
cat (				Ser					Leu					Gly		8183
gct g			Cys					Phe					Arg			8231
ctc a		Leu					Ala					Asp				8279
tgg   Trp   2160	Phe	-	-			Ser					Tyr		-	_		8327
cgt g Arg	_	_		-	Trp		_		_	Leu					Val	8375
GJA .				Tyr					Arg		acgg	<b>J</b> ggag	gct a	aaaca	actcca	8428
															ttttt gtggct	
	ctta	igc o	cctag	gtcad	g g	tago	etgte	, aaa							cagaga	
		.au 1	-9900		JE GI	-aga	Laac	, .								5055
<210:		542														

<211> 8642 <212> DNA

<213> HCV

<220>

```
<221> CDS
<222> (1802)...(8407)
<221> variation
<222> 6268
<223> r = a or g
<221> variation
<222> 4446
<223> r = a or g
<400> 2
accagecece gattggggge gacactecae catagateae teecetgtga ggaactaetg 60
tetteaegea gaaagegtet agecatggeg ttagtatgag tgtegtgeag ceteeaggae 120
cccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaceggg teetttettg gateaacecg etcaatgeet ggagatttgg gegtgeecec 240
gegagactge tageegagta gtgttgggte gegaaaggee ttgtggtaet geetgatagg 300
gtgcttgcqa gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
cggccgcttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct 480
ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgacgggcgt teettgegca getgtgeteg acgttgteac tgaageggga agggactgge 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtateeat catggetgat geaatgegge ggetgeatac gettgateeg getaeetgee 780
cattegacca ccaagegaaa categeateg agegageaeg tacteggatg gaageeggte 840
ttgtcgatca ggatgatctg gacgaagagc atcagggget cgcgccagcc gaactgttcg 900
ccaqqctcaa qqcqcqcatq cccqacqqcq aqqatetcqt cqtqacccat qqcqatqcct 960
gettgeegaa tateatggtg gaaaatggee gettttetgg atteategae tgtggeegge 1020
tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggcggcga atgggctgac cgcttcctcg tgctttacgg tatcgccgct dccgattcgc 1140
agegeatege ettetatege ettettgaeg agttettetg agttegegee eagatgttaa 1200
cagaccacaa cggtttccct ctagcgggat caattccgcc cccccccta acgttactgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
gccgtctttt ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc 1380
taggggtctt tcccctctcg ccaaaggaat gcaaggtctg ttgaatgtcg tgaaggaagc 1440
agttcctctg gaagcttctt gaagacaaac aacgtctgta gcgacccttt gcaggcagcg 1500
qaaccccca cctqqcqaca qqtqcctctq cqqccaaaaq ccacqtqtat aagatacacc 1560
tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggctetcc tcaagegtat tcaacaaggg gctgaaggat gcccagaagg taccccattg 1680
tatgggatet gatetgggge eteggtgeac atgetttaca tgtgtttagt egaggttaaa 1740
aaacgtotag gccccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800
c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849
  Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly
ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg
                                                                  1897
Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg
ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg
                                                                  1945
Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu
        35
caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc
                                                                  1993
Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val
                         55
```

atc c Ile L 65	ete Leu	ctc Leu	acg Thr	tgc Cys	gcg Ala 70	atc Ile	cac His	cca Pro	gag Glu	cta Leu 75	atc Ile	ttt Phe	acc Thr	atc Ile	acc Thr 80	2041
aaa a Lys I																2089
ata a Ile T																2137
tgc a Cys M	Met	ctg Leu 115	gtg Val	cgg Arg	aag Lys	gtt Val	gct Ala 120	eja aaa	ggt Gly	cat His	tat Tyr	gtc Val 125	caa Gln	atg Met	gct Ala	2185
ctc a Leu M	atg Met 130	aag Lys	ttg Leu	gcc Ala	gca Ala	ctg Leu 135	aca Thr	ggt	acg Thr	tac Tyr	gtt Val 140	tat Tyr	gac Asp	cat His	ctc Leu	2233
acc c Thr F 145	cca Pro	ctg Leu	cgg Arg	gac Asp	tgg Trp 150	gcc Ala	cac His	gcg Ala	ggc	cta Leu 155	cga Arg	gac Asp	ctt Leu	gcg Ala	gtg Val 160	2281
gca g Ala V	gtt Val	gag Glu	ccc Pro	gtc Val 165	gtc Val	ttc Phe	tct Ser	gat Asp	atg Met 170	gag Glu	acc Thr	aag Lys	gtt Val	atc Ile 175	acc Thr	2329
tgg g Trp G																2377
gtc t Val S	tcc Ser	gcc Ala 195	cgc Arg	agg Arg	gjå aaa	agg Arg	gag Glu 200	ata Ile	cat His	ctg Leu	gga Gly	ccg Pro 205	gca Ala	gac Asp	agc Ser	2425
ctt g Leu G	gaa	999														
2	210	Gly	Gln	gjå aaa	tgg Trp	cga Arg 215	ctc Leu	ctc Leu	gcg Ala	cct Pro	att Ile 220	acg Thr	gcc Ala	tac Tyr	tcc Ser	2473
caa c Gln G 225	210 cag	Gly acg	Gln cga	ggc	Trp	Arg 215 ctt	Leu ggc	Leu tgc	Ala atc	Pro atc	Ile 220 act	Thr agc	Ala ctc	Tyr	Ser ggc	2473 2521
caa c	cag Sln gac	Gly acg Thr	Gln cga Arg	Gly ggc Gly	cta Leu 230	Arg 215 ctt Leu gag	aaa gla aac ren	tgc Cys gag	Ala atc Ile	atc Ile 235	Ile 220 act Thr	Thr agc Ser	Ala ctc Leu tcc	Tyr aca Thr	ggc Gly 240	
caa c Gln 6 225 cgg g	cag Gln gac Asp	acg Thr agg Arg	cga Arg aac Asn	ggc Gly cag Gln 245	cta Leu 230 gtc Val	Arg 215 ctt Leu gag Glu acc	ben ggc gly tgc	tgc Cys gag Glu gtc	atc Ile gtc Val 250	atc Ile 235 caa Gln	11e 220 act Thr gtg Val	Thr agc Ser gtc Val	Ala ctc Leu tcc Ser	aca Thr acc Thr 255	ggc Gly 240 gca Ala	2521

					aat Asn										2713
					tcc Ser 310										2761
		_			agg Arg	_	_			_		-			2809
	_	_			agc Ser						-			_	2857
-			_		ggt Gly	_				_			~		2905
					gcc Ala										2953
					gag Glu 390										3001
_					tcc Ser		-	-					_		3049
-				-	cct Pro		_		_	_		_		-	3097
					caa Gln										3145
					ggt Gly										3193
					aga Arg 470										3241
					acc Thr										3289
					gac Asp										3337

gac Asp	tcg Ser	acc Thr 515	act Thr	atc Ile	ctg Leu	ggc ggc	atc Ile 520	Gly	aca Thr	gtc Val	ctg Leu	gac Asp 525	caa Gļn	gcg Ala	gag Glu	3385
acg Thr	gct Ala 530	gga Gly	gcg Ala	cga Arg	ctc Leu	gtc Val 535	gtg Val	ctc Leu	gcc Ala	acc Thr	gct Ala 540	acg Thr	cct Pro	ecg Pro	gga Gly	3433
	gtc Val															3481
act Thr	gga Gly	gaa Glu	atc Ile	ccc Pro 565	ttt Phe	tat Tyr	ggc	aaa Lys	gcc Ala 570	atc Ile	ccc Pro	atc Ile	gag Glu	acc Thr 575	atc Ile	3529
	gjå aaa															3577
	ctc Leu															3625
tac Tyr	cgg Arg 610	ggc Gly	ctt Leu	gat Asp	gta Val	tcc Ser 615	gtc Val	ata Ile	cca Pro	act Thr	agc Ser 620	gga Gly	gac Asp	gtc Val	att Ile	3673
	gta Val															3721
	gtg Val															3769
	gac Asp															3817
	tca Ser															3865
att Ile	tac Tyr 690	agg Arg	ttt Phe	gtg Val	act Thr	cca Pro 695	gga Gly	gaa Glu	cgg Arg	ccc Pro	tcg Ser 700	gly	atg Met	ttc Phe	gat Asp	3913
tcc Ser 705	tcg Ser	gtt Val	ctg Leu	tgc Cys	gag Glu 710	tgc Cys	tat Tyr	gac Asp	gcg Ala	ggc Gly 715	tgt Cys	gct Ala	tgg Trp	tac Tyr	gag Glu 720	3961
					, 10											

		_		_	_	_	_		_	gag Glu				_	_	4057
	Thr									ttc Phe						4105
										gca Ala						4153
										tgg Trp 795						4201
-								_		GJA aaa						4249
										act Thr						4297
										gac Asp						4345
										gca Ala						4393
					Ser					ggc Gly					Ser	4441
					870					875					880	
					att					875 gtc Val					ttc	4489
Gly gat	Xaa gag	Pro atg	Ala gaa	Ile 885 gag	att Ile tgc	Pro gcc	Asp tca	Arg	Glu 890 ctc	gtc	Leu tac	Tyr atc	Arg gaa	Glu 895 cag	ttc Phe gga	4489 4537
Gly gat Asp	Xaa gag Glu cag	Pro atg Met	gaa Glu 900 gcc	Ile 885 gag Glu gaa	att Ile tgc Cys	Pro gcc Ala ttc	Asp tca Ser	cac His 905	Glu 890 ctc Leu aag	gtc Val	tac Tyr	Tyr atc Ile	Arg gaa Glu 910 ttg	Glu 895 cag Gln ctg	ttc Phe gga Gly caa	
gat Asp atg Met	Xaa gag Glu cag Gln	atg Met ctc Leu 915	gaa Glu 900 gcc Ala	Ile 885 gag Glu gaa Glu caa	att Ile tgc Cys caa Gln	gcc Ala ttc Phe	tca Ser aaa Lys 920 gct	cac His 905 cag Gln	Clu 890 ctc Leu aag Lys	gtc Val cct Pro	tac Tyr atc Ile	atc Ile 999 Gly 925 gtg	gaa Glu 910 ttg Leu	Glu 895 cag Gln ctg Leu	ttc Phe gga Gly caa Gln	4537

agc ggg ata Ser Gly Ile						4729
gcg ata gca Ala Ile Ala		: Ala Phe T		_	_	4777
acc acc caa Thr Thr Gln 99	His Thr Le					4825
gcc caa ctt Ala Gln Leu 1010	_		-			4873 /
atc gct gga Ile Ala Gly 1025		. Gly Ser I		Gly Lys Val		4921
gat att ttg Asp Ile Leu						4969
ttt aag gtc Phe Lys Val		Glu Met F			Val Asn	5017
cta ctc cct Leu Leu Pro 107	Ala Ile Le			gtc gtc ggg Val Val Gly 1085		5065
tgc gca gcg Cys Ala Ala 1090						5113
cag tgg atg Gln Trp Met 1105		ı Ile Ala P		Arg Gly Asn	_	5161
tcc ccc acg Ser Pro Thr						5209
cag atc ctc Gln Ile Leu	-	Thr Ile T	-		Leu His	5257
cag tgg atc Gln Trp Ile 115	Asn Glu Asp	-	•	tcc ggc tcg Ser Gly Ser 1165		5305

tgg ctc cag tcc a Trp Leu Gln Ser L 1185	ag ctc ctg ccg ys Leu Leu Pro 1190	cga ttg ccg gga Arg Leu Pro Gly 1195	gtc ccc ttc Val Pro Phe	ttc 5401 Phe 1200
tca tgt caa cgt g Ser Cys Gln Arg G 1	ggg tac aag gga ly Tyr Lys Gly .205	gtc tgg cgg ggc Val Trp Arg Gly 1210	gac ggc atc Asp Gly Ile 1215	Met
caa acc acc tgc c Gln Thr Thr Cys P 1220	ca tgt gga gca Pro Cys Gly Ala	cag atc acc gga Gln Ile Thr Gly 1225	cat gtg aaa His Val Lys 1230	aac 5497 Asn
tgt tcc atg agg a Cys Ser Met Arg I 1235	tc gtg ggg cct le Val Gly Pro 124	Arg Thr Cys Ser	aac acg tgg Asn Thr Trp 1245	cat 5545 His
gga aca ttc ccc a Gly Thr Phe Pro I 1250			Cys Thr Pro	
ccg gcg cca aat t Pro Ala Pro Asn T 1265				
tac gtg gag gtt a Tyr Val Glu Val T 1	acg cgg gtg ggg Thr Arg Val Gly .285	gat ttc cac tac Asp Phe His Tyr 1290	gtg acg ggc Val Thr Gly 1295	Met
acc act gac aac g Thr Thr Asp Asn V 1300			-	
ttc aca gaa gtg g Phe Thr Glu Val A 1315	gat ggg gtg cgg Asp Gly Val Arg 132	Leu His Arg Tyr	get eca geg Ala Pro Ala 1325	tgc 5785 Cys
aaa ccc ctc cta c Lys Pro Leu Leu A 1330	egg gag gag gto Arg Glu Glu Val 1335	aca tte ctg gtc Thr Phe Leu Val 134	. Gly Leu Asn	caa 5833 Gln
tac ctg gtt ggg t Tyr Leu Val Gly S 1345				
gtg ctc act tcc a Val Leu Thr Ser M 1				Thr
gct aag cgt agg c Ala Lys Arg Arg L 1380				
tca gct agc cag c Ser Ala Ser Gln L 1395	etg tot gog oot Seu Ser Ala Pro 140	Ser Leu Lys Ala	aca tgc act Thr Cys Thr 1405	acc 6025 Thr

cgt cat gac tcc Arg His Asp Ser 1410		Leu Ile Glu		
cgg cag gag atg Arg Gln Glu Met 1425			Glu Ser Glu	
gta gta att ttg Val Val Ile Leu				
agg gaa gta tcc Arg Glu Val Ser 1460	Val Pro Ala Glu			Lys Phe
cct cga gcg atg Pro Arg Ala Met 1475		Arg Pro Asp		_
ttr gag tee tgg Xaa Glu Ser Trp 1490		Tyr Val Pro		
tgt cca ttg ccg Cys Pro Leu Pro 1505		_	Pro Pro Pro	
aag agg acg gtt Lys Arg Thr Val				
gag ctc gcc aca Glu Leu Ala Thr 1540	Lys Thr Phe Gly	-		Val Asp
agc ggc acg gca Ser Gly Thr Ala 1555		Asp Gln Pro		
gcg gga tcc gac Ala Gly Ser Asp 1570		Ser Ser Met		
gag ccg ggg gat Glu Pro Gly Asp 1585	-		Trp Ser Thr	
gag gag gct agt g Glu Glu Ala Ser			_	
aca ggc gcc ctg Thr Gly Ala Leu . 1620	atc acg cca tgc Ile Thr Pro Cys	get geg gag		ctg ccc 6697 Leu Pro
atc aat gca ctg Ile Asn Ala Leu 1635		Leu Arg His		

gct ac Ala Th	a aca r Thr 50	tct Ser	cgc Arg	agc Ser	gca Ala 1655	Ser	ctg Leu	cgg Arg	cag Gln	aag Lys 1660	Lys	gtc Val	acc Thr	ttt Phe	6793
gac ag Asp Ar 1665					Asp					Asp					6841
atg aa Met Ly				Ser					Lys					Glu	6889
gaa go Glu Al			Leu					Ser					Phe		6937
tat gg Tyr Gl		Lys					Leu					Val			6985
atc cg Ile Ar 17						Leu					Glu				7033
gac ac Asp Th 1745					Lys					Cys					7081
aag gg Lys Gl				Pro					Val					Gly	7129
gtt cg Val Ar			Glu					Tyr					Thr		7177
cct ca Pro Gl	g gcc n Ala 179	Val	atg Met	ggc Gly	tct Ser	tca Ser 1800	Tyr	gga Gly	ttc Phe	caa Gln	tac Tyr 180!	Ser	cct Pro	gga Gly	7225
cag cg Gln Ar 18						Asn					Lys				7273
atg gg Met Gl 1825					Thr					Ser					7321
aat ga Asn As				Glu					Gln					Ala	7369
ccc ga Pro Gl	a gcc u Ala	aga Arg 1860	Gln	gcc Ala	ata Ile	agg Arg	tcg Ser 1869	Leu	aca Thr	gag Glu	cgg Arg	ctt Leu 1870	Tyr	atc Ile	7417

Gly Gly Pro 187		Asn Ser				Tyr 7		7465
tgc cgc gcg Cys Arg Ala 1890			Thr Thr					7513
tgt tac ttg Cys Tyr Leu 1905	Lys Ala				Ala Lys			7561
tgc acg atg Cys Thr Met	_			_	_	Cys C		7609
gcg ggg acc Ala Gly Thr				Leu Arg				7657
atg act aga Met Thr Arg 195	Tyr Ser	Ala Pro				Pro 0		7705
gac ttg gag Asp Leu Glu 1970	_		_			_		7753
gat gca tct Asp Ala Ser 1985	Gly Lys A				Arg Asp			7801
Asp Ala Ser	Gly Lys I	Arg Val ' 1990 gcg tgg (	Tyr Tyr gag aca	Leu Thr 1999 gct aga	Arg Asp	Pro T	Thr Thr 2000 gtc aat	7801 7849
Asp Ala Ser 1985 ccc ctt gcg	cgg gct gArg Ala 2 2005	Arg Val ' 1990 gcg tgg ( Ala Trp (	Tyr Tyr gag aca Glu Thr atg tat	Leu Thr 1995 gct aga Ala Arg 2010 gcg ccc Ala Pro	Arg Asp cac act His Thr	cca g Pro V	Thr Thr 2000 gtc aat Val Asn 2015 gca agg	
Asp Ala Ser 1985 ccc ctt gcg Pro Leu Ala tcc tgg cta	cgg gct gArg Ala A 2005 ggc aac a Gly Asn 12020 atg act comet Thr B	Arg Val 1990  gcg tgg g Ala Trp ( atc	gag aca Glu Thr atg tat Met Tyr 2025	Leu Thr 1995 gct aga Ala Arg 2010 gcg ccc Ala Pro atc ctt	Arg Asp  cac act His Thr  acc ttg Thr Leu  cta gct	cca g Pro N 2 tgg g Trp A 2030 cag g Gin G	Thr Thr 2000 gtc aat Val Asn 2015 gca agg Ala Arg	7849
Asp Ala Ser 1985  ccc ctt gcg Pro Leu Ala  tcc tgg cta Ser Trp Leu  atg atc ctg Met Ile Leu	cgg gct g Arg Ala 2 2005 ggc aac a Gly Asn 1 2020 atg act c Met Thr H	Arg Val 1990  gcg tgg g Ala Trp ( atc	gag aca Glu Thr  atg tat Met Tyr 2025 ttc tcc Phe Ser 2040 cag atc Gln Ile	Leu Thr 1995 gct aga Ala Arg 2010 gcg ccc Ala Pro atc ctt Ile Leu tac ggg	cac act His Thr acc ttg Thr Leu cta gct Leu Ala 2045	cca g Pro V tgg g Trp F 2030 cag g Gln G	Thr Thr 2000 gtc aat Val Asn 2015 gca agg Ala Arg gaa caa Glu Gln	7849 7897
Asp Ala Ser 1985  ccc ctt gcg Pro Leu Ala  tcc tgg cta Ser Trp Leu  atg atc ctg Met Ile Leu 203  ctt gaa aaa Leu Glu Lys	cgg gct gArg Ala A 2005 ggc aac a Gly Asn 1 2020 atg act c Met Thr H	Arg Val 1990  gcg tgg g Ala Trp ( atc atc a Ile Ile I cat ttc i His Phe i 2 Asp Cys ( 2055	gag aca Glu Thr atg tat Met Tyr 2025 ttc tcc Phe Ser 2040 cag atc Gln Ile atc att	gct aga Ala Arg 2010  gcg ccc Ala Pro  atc ctt Ile Leu  tac ggg Tyr Gly  caa cga	acc ttg Thr Leu  cta gct Leu Ala 2045 gcc tgt Ala Cys 2060  ctc cac Leu His	cca g Pro V tgg g Trp A 2030 cag g Gln G	Thr Thr 2000 gtc aat Val Asn 2015 gca agg Ala Arg gaa caa Glu Gln tcc att Ser Ile	7849 7897 7945

tca tgc ctc agg Ser Cys Leu Arg 2100	Lys Leu (		Pro Leu	Arg Val		8137
cgg gcc aga agt Arg Ala Arg Ser 2115			_		Gly Arg Ala	8185
gcc act tgt ggc Ala Thr Cys Gly 2130	Lys Tyr I					8233
aaa ctc act cca Lys Leu Thr Pro 2145	atc ccg of Ile Pro 1 2150	gct gcg tc Ala Ala Se	c cag ttg r Gln Leu 2155	Asp Leu	tcc agc tgg Ser Ser Trp 2160	8281
ttc gtt gct ggt Phe Val Ala Gly						8329
gcc cga ccc cgc Ala Arg Pro Arg 2180	Trp Phe N		Leu Leu	Leu Leu		8377
gta ggc atc tat Val Gly Ile Tyr 2195				ggagct a	aacactcca	8427
ggccaatagg ccatc ttttttttt ttttt ggctccatct tagcc gagagtgctg atact	ttttc tco	cttttttt to cggctagc to	ctctttt gtgaaaggt	ttccttttc	ct ttcctttggt	8547
<210> 3 <211> 2201 <212> PRT <213> HCV						
<220> <221> VARIANT <222> 882 <223> Xaa is Lys	or Arg					
<221> VARIANT <222> 1489 <223> Xaa is Leu						
<400> 3 Met Asp Arg Glu 1		Ala Ser Cys	_	Ala Val I		
	5		10		15	
Leu Ile Leu Leu '	Thr Leu S		туг ьуз			
Leu Ile Trp Trp 3		25		3	30 ,	
Leu Ile Trp Trp	Leu Gln T Pro Pro L	25 Tyr Phe Ile · 40	Thr Arg	3 Ala Glu A 45	30 . Ala His Leu	

Lys	Ile	Leu	Leu	Ala 85	Ile	Leu	Gly	Pro	Leu 90	Met	Val	Leu	Gln	Ala 95	Gly
Ile	Thr	Lys	Val	Pro	Tyr	Phe	Val	Arg 105	Ala	His	Gly	Leu	Ile 110	Arg	Ala
Сув	Met	Leu 115	Val	Arg	Lys	Val	Ala 120	Gly	GΊλ	His	Tyr	Val 125	Gln	Met	Ala
Leu	Met 130	Lys	Leu	Ala	Ala	Leu 135	Thr	Gly	Thr	Tyr	<b>Val</b>	Tyr	Asp	His	Leu
145	Pro				150					155					160
	۷al			165					170					175	
	Gly		180					185					190		
	Ser	195					200					205			
	Glu 210	_		-	_	215					220				
225	Gln Asp				230					235					240
_	Asp			245					250					255	
	His		260					265					270		
_	Gln	275					280					285			
	290 Pro		-			295	-		_		300	_	_		
305	Tyr				310					315					320
	Asp			325					330					335	
_	Gly		340					345					350		
	Ile	355					360					365			
-	370 Phe		_			375	_		_	_	380		_		
385					390					395					400
	Thr			405					410					415	
	His		420					425					430		
	Ala	435					440					445			
	Ala 450				_	455					460				
465	Asp				470					475					480
	Ile			485					490					495	
	Gly		500					505					510		
_	Ser	515				_	520					525			
Thr	Ala 530	Gly	Ala	Arg	Leu	Val 535	Val	Leu	Ala	Thr	Ala 540	Thr	Pro	Pro	Gly

Ser 545	Val	Thr	Val	Pro	His 550	Pro	Asn	Ile	Glu	Glu 555	Val	Ala	Leu	Ser	Ser 560
Thr	Gly	Glu	Ile	Pro 565	Phe	Tyr	Gly	Lys	Ala 570	Ile	Pro	Ile	Glu	Thr 575	Ile
Lys	Gly	Gly	Arg 580	His	Leu	Ile	Phe	Cys 585	His	Ser	Lys	ГÀЗ	Lys 590	Суз	Asp
Glu	Leu	Ala 595	Ala	Lys	Leu	Ser	Gly 600	Leu	Gly	Leu	Asn	Ala 605	Val	Ala	Tyr
Tyr	Arg 610	Gly	Leu	Asp	Val	Ser 615	Val	Ile	Pro	Thr	Ser 620	Gly	Asp	Val	Ile
Val 625	Val	Ala	Thr	Asp	Ala 630	Leu	Met	Thr	Gly	Phe 635	Thr	Gly	Asp	Phe	Asp 640
Ser	Val	Ile	Asp	Cys 645	Asn	Thr	Cys	Val	Thr 650	Gln	Thr	Val	Asp	Phe 655	Ser
	_		660					665				Pro	670	_	
		675			_	_	680	_		-		Gly 685			_
	690					695	_		_		700	Gly			_
Ser 705	Ser	Val	Leu	Cys	Glu 710	Cys	Tyr	Asp	Ala	Gly 715	Cys	Ala	Trp	Tyr	G.Lu 720
Leu	Thr	Pro	Ala	Glu 725	Thr	Ser	Val	Arg	Leu 730	Arg	Ala	Tyr	Leu	Asn 735	Thr
	_		740		_		<del></del>	745				Trp	750		
Phe	Thr	Gly 755	Leu	Thr	His	Ile	Asp 760	Ala	His	Phe	Leu	Ser 765	Gln	Thr	Lys
	770	-	-			775	-				780	Gln			
785					790					795		Gln			800
-			-	805	_				810	-		Thr		815	
-	_		820					825				Thr	830		
		835					840					Glu 845			
	850	_				855	_				860	Leu			-
865				_	870					875		Ile			880
				885					890			Tyr		895	
_			900					905				Ile	910		_
		915					920		_			Gly 925			
	930		_			935					940	Val			-
Trp 945	Arg	Thr	Leu	Glu	Ala 950		Trp	Ala	Lys	His 955	Met	Trp	Asn	Phe	Ile 960
Ser	Gly	Ile	Gln	Tyr 965	Leu	Ala	Gly	Leu	Ser 970	Thr	Leu	Pro	Gly	Asn 975	Pro
			980					985				Thr	990		
Thr	Thr	Gln 995	His	Thr	Leu	Leu	Phe 1000		Ile	Leu	Gly	Gly 1005		Val	Ala

Ala	Gln 1010		Ala	Pro	Pro	Ser 1015		Ala	Ser	Ala	Phe 102		Gly	Ala	Gly
Ile	Ala	Gly	Ala	Ala	Val	Gly	Ser	Ile	Gly	Leu	Gly	Lys	Val	Leu	Val
1025					1030					103					1040
Asp	Ile	Leu	Ala	Gly	Tyr	Gly	Ala	Gly			Gly	Ala	Leu	Val	Ala
				104					1050					1055	
Phe	Lys	Val	Met 1060		Gly	Glu	Met	Pro 1069		Thr	Glu	Asp	Leu 1070		Asn
Leu	Leu	Pro			Leu	Ser	Pro			Leu	Val	Val			Val
		1075					1080					108			
Cys	Ala 1090		Ile	Leu	Arg	Arg 109		Va1	Gly	Pro	Gly 110		Gly	Ala	Val
Gln	Trp	Met	Asn	Arg	Leu	Ile	Ala	Phe	Ala	Ser	Arg	Gly	Asn	His	Val
1105	_				1110					1115					1120
Ser	Pro	Thr	His	Tyr	Val	Pro	Glu	Ser	Asp	Ala	Ala	Ala	Arg	Val	Thr
				112					1130					1135	
Gln	Ile	Leu	Ser	Ser	Leu	Thr	Ile	Thr	Gln	Leu	Leu	Lys	Arg	Leu	His
			1140					1145					1150		
	_	1155	5		_	Сув	1160	)				116	5		
Arg	Asp	Val	$\operatorname{Trp}$	Asp	Trp	Ile	Cys	Thr	Val	Leu	Thr	Asp	Phe	Lys	Thr
	1170					1175					118				-
Trp	Leu	Gln	Ser	Lys	Leu	Leu	Pro	Arg	Leu	Pro	Gly	Val	Pro	Phe	Phe
1185					1190					119					1200
				120	5	Lys			1210	)				121!	5
${ t Gln}$	Thr	Thr	Cys	Pro	Cys	Gly	Ala	Gln	Ile	Thr	Gly	His	Val	Lys	Asn
			1220					122					123		
Cys	Ser	Met	Arg	Ile	Val	Gly			Thr	cys	Ser			$\operatorname{Trp}$	His
		1235					1240			_		124			
Gly			Pro	Ile	Asn	Ala		Thr	Thr	Gly		_	Thr	Pro	Ser
_	1250				<b>a</b>	125			<b></b>	<b>3</b>	126		77 -	~1	a1
		Pro	Asn	Tyr		Arg	Ата	ьeu	тр			ALA	Ala	GIU	
1265		<b>a</b> 1	77 <u>~</u> 7	mla	1270		a1	70 000	Dha	127!		37m 3	mb	C7	1280 Mot
_				128	5	Val			1290	)				129	5
Thr	Thr.	Asp			Lys	Суз	Pro			Val	Pro	Ala			Phe
<b>5</b> 1	cm1	<b>~</b> 7	1300		<b>~</b> 1	77- T	<b>3</b>	130		7	m	37-	131(		G
Pne	ınr			Asp	GIY	Val	-		HIS	Arg	Tyr	132		ALA	Cys
T	Dwo	1315		70 2000	<i>α</i> 1	Glu	1320		Dho	T.011	17-1			λen	Gla
пуs	1330		neu	Arg	GIU	1335		T 11T	FILE	пец	134		пец	LIG11	GLIL
מרג איז			G] w	Ser	Gln	Leu		Cve	GT11	Dro			Den	17a 1	בומ
1345		vaı	GTĀ	Ser	1350		P.T.O	Cys	GIU	1355		FIC	rap	Val	1360
		Фþж	Car	Mot		Thr	7) cm	Dro	Car			Thr	Δ1 =	Glu	
				136	5				1370	)				1375	5
Ala	Lys	Arg	Arg 1380		Ala	Arg	Gly	Ser 138		Pro	Ser		Ala 139		Ser
Ser	Ala	Ser	${\tt Gln}$	Leu	Ser	Ala	${\tt Pro}$	Ser	Leu	Lys	Ala	Thr	Cys	Thr	Thx
		1395					1400					1405			
Arg	His	Asp	Ser	Pro	Asp	Ala	Asp	Leu	Ile	Glu	Ala	Asn	Ļeu	Leu	$\mathtt{Trp}$
	1410					1415					1420				
		Glu	Met	Gly	${\tt Gly}$	Asn	Ile	Thr	Arg			Ser	Glu	Asn	
1425					1430					1435				_	1440
Val	Val	Ile	Leu	Asp 144!		Phe	Glu	Pro	Leu 145(		Ala	Glu	Glu	Asp 1455	
Arg	Glu	Val	Ser 146		Pro	Ala		Ile 1469		Arg	Arg	Ser	Arg 1470		Phe

		147	5				148	0		Asp		148	5		
Xaa	Glu 149		Trp	Lys	Asp	Pro 149		Tyr	Val	Pro	Pro		Val	His	Gly
Cys 150	Pro		Pro	Pro	Ala 1510	Lys		Pro	Pro	Ile 151		Pro	Pro	Arg	Arg 1520
Lys	Arg	Thr	Val	Val 1529		Ser	Glu	Ser	Thr 153	Val 0	Ser	Ser	Ala	Leu 153!	
Glu	Leu	Ala	Thr 154		Thr	Phe	Gly	Ser		Glu	Ser	Ser	Ala 155		Asp
Ser	Gly	Thr 155		Thr	Ala	Ser	Pro	-	Gln	Pro	Ser	Asp 156	_	Gly	Asp
Ala	Gly 1570		Asp	Val	Glu	Ser 157		Ser	Ser	Met	Pro		Leu	Glu	Gly
Glu	Pro	Gly	Asp	Pro	Asp	Leu	Ser	Asp	Gly	Ser	Trp	Ser	Thr	Val	Ser
1589					1590					159					1600
Glu	Glu	Ala	Ser	Glu 160		Val	Val	Cys	Cys 161	Ser O	Met	Ser	Tyr	Thr 161	
			162	0				162	5	Glu			163	)	
Ile	Asn	Ala 163		Ser	Asn	Ser	Leu 164		Arg	His	His	Asn 164		Val	Tyr
Ala	Thr 1650		Ser	Arg	Ser	Ala 165		Leu	Arg	Gln	Lys 1660	_	Val	Thr	Phe
Asp 1665		Leu	Gln	Val	Leu 1670	Asp		His	Tyr	Arg 167!	Asp		Leu	Lys	Glu 1680
		Ala	Lvs	Ala		-	Val	Lvs	Ala	Lys		Leu	Ser	Val	
				1689	5				1690	0				169	õ
			1700	)				170	5	Ala			171	)	
Tyr	GIY	Ala 171		Asp	Val	Arg	Asn 172		Ser	Ser	Lys	Ala 172		Asn	His
Ile			Val	Trp	Lys	Asp	Leu	Leu	Glu	Asp	Thr	Glu	Thr	Pro	<sup>'</sup> Ile
70	1730		-1.	<b>35 - 1</b> -		173		~7	**- 7	m1	1740		~1	-	~7
1745		Thr	iie	Met	1750		Asn	GIU	vaı	Phe 175		vaı	GIN		1760
Lys	Gly	Gly	Arg	Lys 1765		Ala	Arg	Leu	Ile 1770	Val	Phe	Pro	Asp	Leu 1779	
Val	Arg	Val	Cys			Met	Ala	Leu		Asp	Val	Val	Ser		
			1780					178					1790		
Pro	Gln	Ala 179!		Met	Gly	Ser	Ser 180		Gly	Phe	Gln	Tyr 180		Pro	Gly
Gln	Arg 1810		Glu	Phe	Leu	Val 181		Ala	Trp	Lys	Ala 1820		Lys	Cys	Pro
Met	Gly	Phe	Ala	Tyr	Asp	Thr	Arg	Cys	Phe	Asp	Ser	Thr	Val	Thr	Glu
										1835					
Asn	Asp	Ile	Arg	Val 1845		Glu	Ser	Ile	Tyr 1850	Gln )	Cys	Cys	Asp	Leu 1855	
Pro	Glu	Ala	Arg 1860		Ala	Ile	Arg	Ser 186		Ţhr	Glu	Arg	Leu 1870		Ile
Gly	Gly	Pro 1875		Thr	Asn	Ser	Lys 1880		Gln	Asn	Cys	Gly 1885		Arg	Arg
Cys	Arg 1890	Ala		Gly	Val	Leu 189	Thr		Ser	Cys	Gly 1900	Asn		Leu	Thr
Cvs			Lvs	Ala	Ala			Circ	Δτα	Ala			Len	Gln	Asp
1905			-15		1910		Last CI	~ Ys	~+Y	1915		1713	س ب		1920
		Met	Leu	Val			Asp	Asp	Leu	Val		Ile	Cys	Glu	
				1925	;				1020	1				1935	:

```
Ala Gly Thr Gln Glu Asp Glu Ala Ser Leu Arg Ala Phe Thr Glu Ala
                            1945
Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Lys Pro Glu Tyr
                                        1965
       1955
                      1960
Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His
  1970
                    1975 1980
Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr
1985 1990 1995 2000
Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn
             2005 2010 2015
Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro Thr Leu Trp Ala Arg
         2020 2025
                                             2030
Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln
                        2040
                                          2045
       2035
Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile
                     2055
Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser
                 2070
                                  2075
Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala
             2085
                              2090
Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His
               2105 2110
          2100
Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala
      2115 2120
Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu
          2135 2140
   2130
Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp
                2150
                                   2155
Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg
              2165
                              2170 2175
Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly
                          2185
Val Gly Ile Tyr Leu Leu Pro Asn Arg
       2195
<210> 4
<211> 8643
<212> DNA
<213> HCV
<220>
<221> CDS
<222> (1802)...(8407)
<400> 4
accagococo gattgggggc gacactocac catagatcac teccetgtga ggaactaetg 60
tetteacgca gaaagegtet agecatggeg ttagtatgag tgtegtgeag cetecaggae 120
cccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaccggg tectttettg gatcaacccg etcaatgeet ggagatttgg gegtgeecce 240
gegagaetge tageegagta gtgttgggte gegaaaggee ttgtggtaet geetgatagg 300
gtgettgega gtgeeeeggg aggtetegta gaeegtgeae catgageaeg aatectaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
cggccgcttg ggtggagagg ctattcggct atgactgggc gcaacagaca atcggctgct 480
ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg 540
acetgteegg tgeectgaat gaactgeagg acgaggeage geggetateg tggetggeea 600
cgacgggcgt tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtatecat catggetgat geaatgegge ggetgeatac gettgateeg getaeetgee 780
```

cattegacea ceaagegaaa categeateg agegageaeg tacteggatg gaageeggte ttgtegatea ggegecatg ceegacggeg aggatetegt egtgeeagee gaactgtteg gettgeegaa tateatggtg gaaaatggee gettttetgg atteategae tgggeegge ttgggtgtge ggacegetat eaggacatag egttggetae egttggegega atgggetgae egtteeteg tgetttaeeg tategegee egteteteg aggaeatege ettetatege ettettgaeg agttettetg agttegegee eggategee eggetteeteg tgetttaeeg tategegee eaggaeatege ettettgaeg agttettetg agttegegee eaggaeatege ettettgaeg geegggat eaatteegee eeceeeta aegttaetgg eeggetettt ggeaatgtga ggegeeggaa acetggeeet gtettettga egggeattee taggggtett teceeteeg eaaaggaat geaaggtetg ttgaatgteg tgaaggaage agtteetetg gaagettett gaagaeaae aaegtetgta gegaeeettt geaggeageg gaaeeeecee eagtgeeeet teaageggat teaaeaaggeg geaeaaaee eagtgeeet tgaaggaagge aaaggeeet teaaagggat teaaeaaggg getgaaggat geeeagaagg taceeeattg tatgggatet gatetggge eteggtgeae atgetttaea tgtgtttagt egaggttaga aaaegtetag geeeeeegaa eeggggee gtggttteee ttgaaaaae eeggtgae eteggtgeae atgetttaea tgtgtttagt egaggttaga aaaegtetag geeeeeegaa eeggggae gtggttttee ttgaaaaae eegggtae eteggtgeae atgetttaea tgtgtttagt egaggttaga aaaegtetag geeeeeegaa eeggggae gtggttttee ttgaaaaae aegataatae eaga gae egg gag atg gea gea geg geg gtt tte gta ggt Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly	900 960 1020 1140 1200 1260 1320 1440 1560 1620 1680 1740 1800
ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg 20 25 30	1897
ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu 35 40 45	1945
caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val 50 55 60	1993
atc ctc ctc acg tgc gcg atc cac cca gag cta atc ttt acc atc acc  Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile Thr  65 70 75 80	2041
aaa atc ttg ctc gcc ata ctc ggt cca ctc atg gtg ctc cag gct ggt Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala Gly 85 90 95	2089
ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt gca Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg Ala 100 105 110	2137
tgc atg ctg gtg cgg aag gtt gct ggg ggt cat tat gtc caa atg gct Cys Met Leu Val Arg Lys Val Ala Gly Gly His Tyr Val Gln Met Ala 115 120 125	2185
ctc atg aag ttg gcc gca ctg aca ggt acg tac gtt tat gac cat ctc Leu Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His Leu 130 135 140	2233
acc cca ctg cgg gac tgg gcc cac gcg ggc cta cga gac ctt gcg gtg Thr Pro Leu Arg Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val 145 150 150 160	2281

Āla	_			-		ttc Phe		_	_			_				2329
						gcg Ala	_									2377
						agg Arg										2425
	_		_			cga Arg 215						_	_			2473
	-	_	_			ctt Leu		-				_				2521
						gag Glu										2569
				_		acc Thr		_				_			_	2617
						aag Lys										2665
		275		•			280					285				
acc	caa	275 atg	tac	acc	aat	gtg Val 295	280 gac	cag	gac	ctc	gtc	285 ggc	tgg	caa	gcg	2713
acc Thr	caa Gln 290 ccc	atg Met	tac Tyr gcg	acc Thr	aat Asn tcc	gtg Val	280 gac Asp	cag Gln cca	gac Asp tgc	ctc Leu acc	gtc Val 300 tgc	285 ggc Gly ggc	tgg Trp	caa Gln tcg	gcg Ala gac	2713 2761
acc Thr ccc Pro 305	caa Gln 290 ccc Pro	atg Met ggg Gly	tac Tyr gcg Ala	acc Thr cgt Arg	aat Asn tcc Ser 310	gtg Val 295 ttg	gac Asp aca Thr	cag Gln cca Pro	gac Asp tgc Cys	ctc Leu acc Thr 315	gtc Val 300 tgc Cys	285 ggc Gly ggc Gly	tgg Trp agc Ser	caa Gln tcg Ser	gcg Ala gac Asp 320	
acc Thr ccc Pro 305 ctt Leu	caa Gln 290 ccc Pro tac Tyr	atg Met ggg Gly ttg Leu	tac Tyr geg Ala gtc Val	acc Thr cgt Arg acg Thr 325	aat Asn tcc Ser 310 aag Lys	gtg Val 295 ttg Leu	gac Asp aca Thr gcc Ala	cag Gln cca Pro gat Asp	gac Asp tgc Cys gtc Val 330 ccc	ctc Leu acc Thr 315 att Ile	gtc Val 300 tgc Cys ccg Pro	285 ggc Gly ggc Gly gtg Val	tgg Trp agc Ser cgc Arg	caa Gln tcg Ser cgg Arg 335	gcg Ala gac Asp 320 cgg Arg	2761
acc Thr ccc Pro 305 ctt Leu ggc Gly	caa Gln 290 ccc Pro tac Tyr gac Asp	atg Met ggg Gly ttg Leu agc Ser	tac Tyr geg Ala gtc Val agg Arg 340 teg	acc Thr cgt Arg acg Thr 325 ggg Gly	aat Asn tcc Ser 310 aag Lys agc	gtg Val 295 ttg Leu cat His	gac Asp aca Thr gcc Ala ctc Leu ctg	cag Gln cca Pro gat Asp tcc Ser 345	gac Asp tgc Cys gtc Val 330 ccc Pro	ctc Leu acc Thr 315 att Ile cgg Arg	gtc Val 300 tgc Cys ccg Pro	ggc Gly ggc Gly gtg Val gtc Val	tgg Trp agc ser cgc Arg tcc ser 350	caa Gln tcg Ser cgg Arg 335 tac Tyr	gcg Ala gac Asp 320 cgg Arg ttg Leu	2761 2809

-		-	ccc Pro	_							_			_	_	3001
		-	aac Asn	-			-	-	_	_						3049
			cac His 420													3097
			gca Ala													3145
			acc Thr													3193
			aac Asn													3241
			tac Tyr													3289
			gcc		-				_	gat Asp		_				3337
	<i></i>	CLY	500	-2				505	-				510			
_	tcg	acc		atc	ctg	ggc		ggc		-	_		caa	gcg	gag	3385
Asp	tcg ser	acc Thr 515 gga	500 act	atc Ile cga	ctg Leu ctc	ggc gtc	Ile 520 gtg	ggc Gly	Thr gcc	Val	Leu gct	Asp 525 acg	caa Gln cct	gcg Ala ccg	gag Glu gga	3385 3433
Asp acg Thr	tcg Ser gct Ala 530	acc Thr 515 gga Gly	500 act Thr	atc Ile cga Arg	ctg Leu ctc Leu	ggc Gly gtc Val 535	Ile 520 gtg Val	ggc Gly ctc Leu	Thr gcc Ala gag	Val acc Thr	gct Ala 540 gtg	Asp 525 acg Thr	caa Gln cct Pro	gcg Ala ccg Pro	gag Glu gga Gly	
acg Thr tcg Ser 545	tcg ser gct Ala 530 gtc Val	acc Thr 515 gga Gly acc Thr	500 act Thr gcg Ala	atc Ile cga Arg cca Pro	ctg Leu ctc Leu cat His 550	ggc Gly gtc Val 535 cca Pro	Ile 520 gtg Val aac Asn	ggc Gly ctc Leu atc Ile	Thr gcc Ala gag Glu	val acc Thr gag Glu 555 atc	gct Ala 540 gtg Val	Asp 525 acg Thr gct Ala	caa Gln cct Pro ctg Leu	gcg Ala ccg Pro tcc Ser	gag Glu gga Gly agc Ser 560	3433
acg Thr tcg Ser 545 act Thr	tcg ser gct Ala 530 gtc Val gga Gly	acc Thr 515 gga Gly acc Thr gaa Glu	500 act Thr gcg Ala gtg Val	atc Ile cga Arg cca Pro ccc Pro 565 cac	ctg Leu ctc Leu cat His 550 ttt Phe ctc	ggc Gly gtc Val 535 cca Pro tat Tyr	Ile 520 gtg Val aac Asn ggc Gly	ggc Gly ctc Leu atc Ile aaa Lys	Thr  gcc Ala  gag Glu  gcc Ala 570 cat	Val acc Thr gag Glu 555 atc Ile	gct Ala 540 gtg Val ccc Pro	Asp 525 acg Thr gct Ala atc Ile	caa Gln cct Pro ctg Leu gag Glu	gcg Ala ccg Pro tcc Ser acc Thr 575	gag Glu gga Gly agc Ser 560 atc Ile	3433 3481

			cca act agc Pro Thr Ser 620		
			ggc ttt acc Gly Phe Thr 635		
tca gtg atc Ser Val Ile	gac tgc aat Asp.Cys Asn 645	aca tgt gtc Thr Cys Val	acc cag aca Thr Gln Thr 650	gtc gac ttc Val Asp Phe 655	agc 3769 Ser
			acg acc gtg Thr Thr Val		
			act ggt agg Thr Gly Arg		
~~			cgg ccc tcg Arg Pro Ser 700		-
			gcg ggc tgt Ala Gly Cys 715		
			ttg cgg gct Leu Arg Ala 730		Thr
			ctg gag ttc Leu Glu Phe		
			cat ttc ttg His Phe Leu		
			gta gca tac Val Ala Tyr 780		
			teg tgg gac Ser Trp Asp 795		
tgt ctc ata Cys Leu Ile	cgg cta aag Arg Leu Lys 805	cct acg ctg Pro Thr Leu	cac ggg cca His Gly Pro 810	acg ccc ctg Thr Pro Leu 815	Leu
tat agg ctg Tyr Arg Leu	gga gcc gtt Gly Ala Val 820	caa aac gag Gln Asn Glu 825	gtt act acc Val Thr Thr	aca cac ccc Thr His Pro 830	ata 4297 Ile

acc aaa tac at Thr Lys Tyr II 835				u Val Val	
agc acc tgg gt Ser Thr Trp Va 850		Gly Val Le			
tgc ctg aca ac Cys Leu Thr Th 865					
gga agg ccg go Gly Arg Pro Al			lu Val Leu Ty		
gat gag atg ga Asp Glu Met Gl 90					
atg cag ctc go Met Gln Leu Al 915				y Leu Leu	
aca gcc acc as Thr Ala Thr Ly 930		Ala Ala Al			
tgg cgg acc ct Trp Arg Thr Le 945					_
agc ggg ata ca Ser Gly Ile Gl	-		er Thr Leu Pr		
gcg ata gca to Ala Ile Ala Se 98	r Leu Met Ala	-			
acc acc caa ca Thr Thr Gln Hi 995				y Trp Val	
gcc caa ctt gc Ala Gln Leu Al 1010		Ala Ala Se	-		
atc gct gga gc Ile Ala Gly Al 1025					
gat att ttg go Asp Ile Leu Al		Ala Gly Va			Ala

ttt aag gt Phe Lys Va				Ser Thr			
cta ctc cc Leu Leu Pr 10	Ala Ile					Gly Val	
tgc gca gc Cys Ala Al 1090			His Val				
cag tgg at Gln Trp Me 1105	g aac cgg : Asn Arg	ctg ata Leu Ile 1110	gcg ttc Ala Phe	gct tcg Ala Ser 111	Arg Gly	aac cac Asn His	gtc 5161 Val 1120
tcc ccc ac Ser Pro Th		Val Pro					Thr
cag atc ct Gln Ile Le				Gln Leu			
cag tgg at Gln Trp Il 11	e Asn Glu					Ser Trp	
aga gat gt Arg Asp Va 1170	t tgg gat l Trp Asp	tgg ata Trp Ile 117	Cys Thr	gtg ttg Val Leu	act gat Thr Asp 1180	ttc aag Phe Lys	gcc 5353 Ala
tgg ctc ca Trp Leu Gl 1185					Gly Val		
tca tgt ca Ser Cys Gl		Tyr Lys					Met
caa acc ac Gln Thr Th				Ile Thr			
tgt tcc at Cys Ser Me 12	t Arg Ile					Thr Trp	
gga aca tt Gly Thr Ph 1250	c ccc att e Pro Ile	aac gcg Asn Ala 125	Tyr Thr	acg ggc Thr Gly	ccc tgc Pro Cys 1260	acg ccc Thr Pro	tcc 5593 Ser
ccg gcg cc Pro Ala Pr 1265					Val Ala		

	gtg Val				Arg					His					Met	5689
	act Thr			Val					Gln					Glu		5737
	aca Thr		Val					Leu					Pro			5785
	CCC Pro 1330	Leu					Val					Gly				5833
	ctg Leu 5					Leu					Glu					5881
	ctc Leu				Leu					His					Thr	5929
	aag Lys			Leu					Pro					Ser		5977
	gct Ala		Gln					Ser					Cys			6025
	cat His 1410	Asp					Asp					Asn				6073
_	cag Gln		-			Asn			-		Glu		_		_	6121
	gta Val				Ser					Gln					Glu	6169
agg Arg	gaa Glu	gta Val	tcc Ser 1460	Val	ccg Pro	gcg Ala	gag Glu	atc Ile 1465	Leu	cgg Arg	agg Arg	tcc Ser	agg Arg 1470	Lys	ttc Phe	6217
	cga Arg		Met					Arg					Pro		J	6265
	gag Glu 1490	Ser					Āsp					Val			Gly 999	6313

tgt cca ttg Cys Pro Leu 1505					Pro Pro			6361
aag agg acg Lys Arg Thr	_	Leu Ser				Ala		6409
gag ctc gcc Glu Leu Ala				Ser Glu				6457
agc ggc acg Ser Gly Thr 155!	Ala Thr					Asp		6505
gcg gga tcc Ala Gly Ser 1570			Tyr Ser	_				6553
gag ccg ggg Glu Pro Gly 1585					Trp Ser			6601
gag gag gct Glu Glu Ala		Asp Val				Tyr :		6649
acg ggc gcc Thr Gly Ala				Ala Glu				6697
atc aat gca Ile Asn Ala 1635	Leu Ser					Leu V		6745
gct aca aca Ala Thr Thr 1650	_		Ser Leu					<sup>-</sup> 6793
gac aga ctg Asp Arg Leu 1665	Gln Val				Asp Val			6841
atg aag gcg Met Lys Ala	aag gcg			-	ctt cta	_		6889
	Lys Ala . 1685		Val Lys	Ala Lys 1690	Leu Leu		vai Giu 1695	
gaa gcc tgt Glu Ala Cys	1685 aag ctg	acg ccc	cca cat	1690 tcg gcc Ser Ala	aga tct	aaa t	1695 tt ggc	6937

1730	tg tgg aag al Trp Lys	gac ttg Asp Leu 1735	ctg gaa Leu Glu	gac act ga Asp Thr Gl 1740	g aca cca u Thr Pro	att 7033 Ile
gac acc acc at Asp Thr Thr II 1745	tc atg gca le Met Ala 175	Lys Asn	gag gtt Glu Val	ttc tgc gt Phe Cys Va 1755	c caa cca 1 Gln Pro	gag 7081 Glu 1760
aag ggg ggc c	gc aag cca rg Lys Pro 1765	gct cgc Ala Arg	ctt atc Leu Ile 1770	Val Phe Pr	a gat ttg o Asp Leu 1775	Gly
gtt cgt gtg t Val Arg Val C	gc gag aaa ys Glu Lys .780	atg gcc Met Ala	ctt tac Leu Tyr 1785	gat gtg gt Asp Val Va	c tcc acc l Ser Thr 1790	ctc 7177 Leu
cct cag gcc g Pro Gln Ala V 1795	rtg atg ggo Val Met Gly	tct tca Ser Ser 180	Tyr Gly	Phe Gln Ty	c tot cot r Ser Pro	gga 7225 Gly
cag cgg gtc g Gln Arg Val G 1810	gag ttc cto	g gtg aat 1 Val Asn 1815	gcc tgg Ala Trp	aaa gcg aa Lys Ala Ly 1820	g aaa tgc rs Lys Cys	cct 7273 Pro
atg ggc ttc g Met Gly Phe A 1825	gca tat gad Ala Tyr Ası 18:	Thr Arg	tgt ttt Cys Phe	gac tca ac Asp Ser Th 1835	g gtc act r Val Thr	gag 7321 Glu 1840
aat gac atc c Asn Asp Ile A	gt gtt gag Arg Val Glo 1845	g gag tca ı Glu Ser	atc tac Ile Tyr 1850	Gln Cys Cy	rt gac ttg rs Asp Leu 185!	Ala
ccc gaa gcc a Pro Glu Ala A 1	aga cag gco Arg Gln Ala 1860	ata agg a Ile Arg	tcg ctc Ser Leu 1865	aca gag cg Thr Glu Ar	g ctt tac g Leu Tyr 1870	atc 7417 Ile
Pro Glu Ala A	Arg Gln Ala .860 stg act aa	a Ile Arg	Ser Leu 1865 ggg cag Gly Gln	aac tgc gg Asn Cys Gl	rg Leu Tyr 1870 gc tat cgc	Ile cgg 7465
Pro Glu Ala A 1 ggg ggc ccc c Gly Gly Pro L	arg Gln Ala .860 etg act aa .eu Thr Ass	tct aaa Ser Lys 188	Ser Leu 1865 ggg cag Gly Gln 0	aac tgc gg Asn Cys Gl 19	rg Leu Tyr 1870 gc tat cgc y Tyr Arg 885	cgg 7465 Arg aca 7513
Pro Glu Ala A  ggg ggc ccc c Gly Gly Pro L 1875  tgc cgc gcg a Cys Arg Ala S	arg Gln Ala 860 etg act aa Geu Thr Ass agc ggt gt Ger Gly Va	tct aaa Ser Lys 188 ctg acg Leu Thr 1895 c gcg gcc	ser Leu 1865  ggg cag Gly Gln 0  acc agc Thr Ser	aac tgc gg Asn Cys Gl tgc ggt aa Cys Gly As 1900 gct gcg aa	g Leu Tyr 1870 gc tat cgc y Tyr Arg 885 at acc ctc in Thr Leu ag ctc cag	cgg 7465 Arg 7513 Thr 7561
Pro Glu Ala A  ggg ggc ccc c Gly Gly Pro L  1875  tgc cgc gcg a Cys Arg Ala S  1890  tgt tac ttg a Cys Tyr Leu L	arg Gln Ala 860  etg act aa eu Thr Asi agc ggt gta Ser Gly Va aag gcc gc Lys Ala Ala 19	tct aaa Ser Lys 188 ctg acg Leu Thr 1895 cgcg gcc Ala Ala	ggg cag Gly Gln  acc agc Thr Ser  tgt cga Cys Arg	aac tgc gg Asn Cys Gl tgc ggt aa Cys Gly As 1900 gct gcg aa Ala Ala Ly 1915 gtc gtt at Val Val II	g Leu Tyr 1870 gc tat cgc y Tyr Arg 885 at acc ctc m Thr Leu ag ctc cag ss Leu Gln	cgg 7465 Arg 7513 Thr 7561 Asp 1920 agc 7609 Ser 7609

atg act aga to Met Thr Arg T 1955						7705
gac ttg gag t Asp Leu Glu L 1970	eu Ile Thr S		Ser Asn Va			7753
gat gca tct g Asp Ala Ser G 1985						7801
ccc ctt gcg c Pro Leu Ala A						7849
tcc tgg cta gg Ser Trp Leu G			Ala Pro Th		Ala Arg	7897
atg atc ctg at Met Ile Leu Me 2035					-	7945
ctt gaa aaa go Leu Glu Lys Ai 2050	a Leu Asp C			a Cys Tyr		7993
gag cca ctt ga Glu Pro Leu Aa 2065						3041
gca ttt tca ci Ala Phe Ser Le						3089
tca tgc ctc ag Ser Cys Leu Ar 23			Pro Leu Ar		Arg His	3137
cgg gcc aga ag Arg Ala Arg Se 2115						3185
gec act tgt ge	c aag tac c	tc ttc aac	tgg gca gt	a agg acc	aag ctc 8	3233
2130		eu Phe Asn 135	Trp Ala Va		туя цец	
	2 a atc ccg g	135 et geg tee	cag ttg ga	40 t tta tcc	agc tgg 8	281

```
gee ega eee ege tgg tte atg tgg tge eta etc eta ett tet gta ggg
                                                                8377
Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly
           2180
                               2185
gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca
                                                                8427
Val Gly Ile Tyr Leu Leu Pro Asn Arg *
       2195
                           2200
Etttttttt tttttttt ttttctttt tcccaattt tttccttttc tttcctttgg 8547
tggctccatc ttagccctag tcacggctag ctgtgaaagg tccgtgagcc gcttgactgc 8607
agagagtgct gatactggcc tctctgcaga tcaagt
<210> 5
<211> 8648
<212> DNA
<213> HCV
<220>
<221> CDS
<222> (1802)...(8407)
<400> 5
qccaqccccc gattgggggc gacactccac catagatcac tcccctgtga ggaactactg 60
tetteacgca gaaagegtet agecatggeg ttagtatgag tgtegtgeag cetecaggae 120
cccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
qacqaccqqq teetttettq gatcaacccq ctcaatgcct ggagatttgg gcgtgcccc 240
gegagaetge tageegagta gtgttgggte gegaaaggee ttgtggtaet geetgatagg 300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
cggccgcttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct 480
ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cqacqqqcqt tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaaqtgccq qqqcaqqatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtatecat catggetgat geaatgegge ggetgeatae gettgateeg getaeetgee 780
cattcgacca ccaagcgaaa catcgcatcg agcgagcacg tactcggatg gaagccggtc 840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900
ccaggeteaa ggegegeatg cccgacggeg aggatetegt cgtgacccat ggegatgeet 960
gettgeegaa tateatggtg gaaaatggee gettttetgg atteategae tgtggeegge 1020
tqqqtqtqqc qgaccqctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggcggcga atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc 1140
agogcatogo ottotatogo ottottgaog agttottotg agttogogoo cagatgttaa 1200
cagaccacaa eggttteeet etagegggat caatteegee ecceeceta aegttaetgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
gccgtctttt ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc 1380
taggggtctt tcccctctcg ccaaaggaat gcaaggtctg ttgaatgtcg tgaaggaagc 1440
agtteetetg gaagettett gaagacaaac aacgtetgta gegaceettt geaggeageg 1500
gaacccccca cctqqcqaca qqtqcctctq cggccaaaag ccacgtgtat aagatacacc 1560
tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggetetee teaagegtat teaacaaggg getgaaggat geecagaagg taccecattg 1680
tatgggatet gatetgggge eteggtgeae atgetttaea tgtgtttagt egaggttaaa 1740
aaacgtctag gccccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800
c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849
  Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly
```

_			_		_		_	cac His 25			_			-		1897
								atc Ile			_		-		_	1945
								gtt Val								1993
			_	_				cca Pro	~			•				2041
								cca Pro								2089
								cgc Arg 105								2137
Cys	Met	Leu 115	Val	Arg	Lys	Val	Ala 120	gjà aaa	Gly	His	Tyr	Val 125	Gln	Met	Ala	2185
Leu	Met 130	Lys	Leu	Ala	Ala	Leu 135	Thr	ggt Gly	Thr	Tyr	Val 140	Tyr	Asp	His	Leu	2233
		ctg	cqq	gac		gcc	cac	gcq.	ggc.	cta	cga.	gac	ctt	gcg.	ata .	2281
145	Pro			Asp	Trp 150	Ala	His	Ala			Arg		Leu	Ala		
145 gca	gtt	Leu gag	Arg ccc	gtc	150 gtc	ttc	tct		Gly atg	Leu 155 gag	acc	Asp	gtt	atc	Val 160 acc	2329
gca Ala tgg	gtt Val ggg	Leu gag Glu gca	Arg ccc Pro	gtc Val 165	gtc Val	ttc Phe gcg	tct Ser	Ala gat	atg Met 170	Leu 155 gag Glu atc	acc Thr	Asp aag Lys ttg	gtt Val ggc	atc Ile 175 ctg	Val 160 acc Thr	
gca Ala tgg Trp	gtt Val ggg Gly	gag Glu gca Ala	ccc Pro gac Asp 180	gtc Val 165 acc Thr	gtc Val gcg Ala	ttc Phe gcg Ala agg	tct Ser tgt Cys	Ala gat Asp ggg	atg Met 170 gac Asp	Leu 155 gag Glu atc Ile	acc Thr atc Ile	aag Lys ttg Leu	gtt Val ggc Gly 190 gca	atc Ile 175 ctg Leu	Val 160 acc Thr	2329
gca Ala tgg Trp gtc Val	gtt Val ggg Gly tcc ser	gag Glu gca Ala gcc Ala 195	ccc Pro gac Asp 180 cgc Arg	gtc Val 165 acc Thr agg Arg	gtc Val gcg Ala 999 Gly	ttc Phe gcg Ala agg Arg	tct Ser tgt Cys gag Glu 200	Ala gat Asp 999 Gly 185 ata	atg Met 170 gac Asp cat His	Leu 155 gag Glu atc Ile ctg Leu	acc Thr atc Ile gga Gly	aag Lys ttg Leu ccg Pro 205	gtt Val ggc Gly 190 gca Ala	atc Ile 175 ctg Leu gac Asp	Val 160 acc Thr ccc Pro agc Ser	2329 2377

cgg Arg	gac Asp	agg Arg	aac Asn	Gln 245	Val	gag Glu	gly ggg	gag Glu	gto Val 250	Gln	gtg Val	gto Val	tcc Ser	acc Thr 255	gca Ala	2569
aca Thr	caa Gln	tct Ser	ttc Phe 260	Leu	gcg Ala	acc Thr	tgo Cys	gtc Val 265	Asn	ggc	gtg Val	tgt Cys	tgg Trp 270	Thr	gtc Val	2617
tat Tyr	cat His	ggt Gly 275	Ala	Gly	tca Ser	aag Lys	acc Thr 280	Leu	gcc Ala	ggc	cca Pro	aag Lys 285	Gly	cca Pro	atc Ile	2665
acc Thr	caa Gln 290	atg Met	tac Tyr	acc Thr	aat Asn	gtg Val 295	gac Asp	cag Gln	gac Asp	ctc Leu	gto Val 300	Gly	tgg Trp	caa Gln	gcg Ala	2713
ccc Pro 305	Pro	gjà aaa	gcg Ala	cgt Arg	tcc Ser 310	ttg Leu	aca Thr	cca Pro	tgc Cys	acc Thr 315	tgc Cys	ggc Gly	agc Ser	tcg Ser	gac Asp 320	2761
ctt Leu	tac Tyr	ttg Leu	gtc Val	acg Thr 325	agg Arg	cat His	gcc Ala	gat Asp	gtc Val 330	att Ile	ccg Pro	gtg Val	cgc Arg	cgg Arg 335	cgg Arg	2809
ggc Gly	gac Asp	agc Ser	agg Arg 340	gly aaa	agc Ser	cta Leu	ctc Leu	tcc Ser 345	ccc Pro	agg Arg	ccc	gtc Val	tcc Ser 350	tac Tyr	ttg Leu	2857
aag Lys	ggc	tct Ser 355	tcg Ser	ggc	ggt Gly	cca Pro	ctg Leu 360	ctc Leu	tgc Cys	ccc Pro	tcg Ser	365 ggg	cac His	gct Ala	gtg Val	2905
ggc	atc Ile 370	ttt Phe	cgg Arg	gct Ala	gcc Ala	gtg Val 375	tgc Cys	acc Thr	cgg Arg	gly	gtt Val 380	gcg Ala	aag Lys	gcg Ala	gtg Val	2953
gac Asp 385	ttt Phe	gta Val	ccc Pro	gtc Val	gag Glu 390	tct Ser	atg Met	gga Gly	acc Thr	act Thr 395	atg Met	cgg Arg	tcc Ser	ccg Pro	gtc Val 400	3001
ttc Phe	acg Thr	gac Asp	aac Asn	tcg Ser 405	tcc Ser	cct Pro	ccg Pro	gcc Ala	gta Val 410	ccg Pro	cag Gln	aca Thr	ttc Phe	cag Gln 415	gtg Val	3049
gcc Ala	cat His	cta Leu	cac His 420	gcc Ala	cct Pro	act Thr	ggt Gly	agc Ser 425	ggc Gly	aag Lys	agc Ser	act Thr	aag Lys 430	gtg Val	ccg Pro	3097
gct Ala	gcg Ala	tat Tyr 435	gca Ala	gcc Ala	caa Gln	gly 999	tat Tyr 440	aag Lys	gtg Val	ctt Leu	gtc Val	ctg Leu 445	aac Asn	ccg Pro	tcc Ser	3145
gtc Val	gcc Ala 450	gcc Ala	acc Thr	cta Leu	ggt Gly	ttc Phe 455	gly āāā	gcg Ala	tat Tyr	atg Met	tct Ser 460	aag Lys	gca Ala	cat <b>His</b>	ggt Gly	3193

																2041
atc Ile 465	gac Asp	cct Pro	aac Asn	atc Ile	aga Arg 470	acc Thr	gly aaa	gta Val	agg Arg	acc Thr 475	Ile	acc Thr	acg Thr	Gly	gcc Ala 480	3241
ccc Pro	atc Ile	acg Thr	tac Tyr	tcc Ser 485	acc Thr	tat Tyr	ggc Gly	aag Lys	ttt Phe 490	ctt Leu	gcc Ala	gac Asp	ggt Gly	ggt Gly 495	tgc Cys	3289
tct Ser	gly aaa	ggc Gly	gcc Ala 500	tat Tyr	gac Asp	atc Ile	ata Ile	ata Ile 505	tgt Cys	gat Asp	gag Glu	tgc Cys	cac His 510	tca Ser	act Thr	3337
gac Asp	tcg Ser	acc Thr 515	act Thr	atc Ile	ctg Leu	gly ggc	atc Ile 520	ggc Gly	aca Thr	gtc Val	ctg Leu	gac Asp 525	caa Gln	gcg Ala	gag Glu	3385
acg Thr	gct Ala 530	gga Gly	gcg Ala	cga Arg	ctc Leu	gtc Val 535	gtg Val	ctc Leu	gcc Ala	acc Thr	gct Ala 540	acg Thr	cct Pro	ccg Pro	gga gga	3433
tcg Ser 545	gtc Val	acc Thr	gtg Val	cca Pro	cat His 550	cca Pro	aac Asn	atc Ile	gag Glu	gag Glu 555	gtg Val	gct Ala	ctg Leu	tcc Ser	agc Ser 560	3481
act Thr	gga Gly	gaa Glu	atc Ile	ccc Pro 565	ttt Phe	tat Tyr	ggc Gly	aaa Lys	gcc Ala 570	atc Ile	ccc Pro	atc Ile	gag Glu	acc Thr 575	atc Ile	3529
aag Lys	eja aaa	eja aaa	agg Arg 580	cac His	ctc Leu	att Ile	ttc Phe	tgc Cys 585	cat His	tcc Ser	aag Lys	aag Lys	aaa Lys 590	tgt Cys	gat Asp	3577
gag Glu	ctc Leu	gcc Ala 595	gcg Ala	aag Lys	ctg Leu	tcc Sér	ggc Gly 600	ctc Leu	gga Gly	ctc Leu	aat Asn	gct Ala 605	gta Val	gca Ala	tat Tyr	3625
tac Tyr	cgg Arg 610	ggc	ctt Leu	gat Asp	gta Val	tcc Ser 615	gtc Val	ata Ile	cca Pro	act Thr	agc Ser 620	gga Gly	gac Asp	gtc Val	att Ile	3673
gtc Val 625	Val	gca Ala	acg Thr	gac Asp	gct Ala 630	cta Leu	atg Met	acg Thr	gly ggc	ttt Phe 635	acc Thr	ggc	gat Asp	ttc Phe	gac Asp 640	3721
tca Ser	gtg Val	atc Ile	gac Asp	tgc Cys 645	aat Asn	aca Thr	tgt Cys	gtc Val	açc Thr 650	cag Gln	aca Thr	gtc Val	gac Asp	ttc Phe 655	Ser	3769
ctg Leu	gac Asp	ccg Pro	acc Thr 660	Phe	acc Thr	att Ile	gag Glu	acg Thr 665	Thr	acc Thr	gtg Val	cca Pro	caa Gln 670	Asp	gcg Ala	3817
gtg Val	tca Ser	cgc Arg 675	Ser	cag Gln	cgg Arg	cga Arg	Gly 680	Arg	act Thr	ggt	agg Arg	ggc Gly 685	Arg	atg Met	Gly	3865

att tac agg Ile Tyr Arg 690	ttt gtg ac	cca gga Pro Gly 695	gaa cgg Glu Arg	ccc tcg Pro Ser 700	ggc atg Gly Met	ttc gat Phe Asp	3913
tcc tcg gtt Ser Ser Val 705	ctg tgc ga Leu Cys Gl 71	ı Cys Tyr	gac gcg Asp Ala	ggc tgt Gly Cys 715	gct tgg Ala Trp	tac gag Tyr Glu 720	3961
ctc acg ccc Leu Thr Pro							4009
cca ggg ttg Pro Gly Leu		s Gln Asp					4057
ttt aca ggc Phe Thr Gly 755							4105
cag gca gga Gln Ala Gly 770							4153
tgc gcc agg Cys Ala Arg 785		a Pro Pro					4201
tgt ctc ata Cys Leu Ile							4249
tat agg ctg Tyr Arg Leu	gga gcc gt Gly Ala Va 820	l Gln Asn	gag gtt Glu Val 825	act acc Thr Thr	aca cac Thr His 830	ccc ata Pro Ile	4297
acc aaa tac Thr Lys Tyr 835	atc atg go Ile Met Al	a tgc atg a Cys Met 840	tcg gct Ser Ala	gac ctg Asp Leu	gag gtc Glu Val 845	gtc acg Val Thr	4345
agc acc tgg Ser Thr Trp 850	gtg ctg gt Val Leu Va	a ggc gga l Gly Gly 855	gtc cta Val Leu	gca gct Ala Ala 860	ctg gcc Leu Ala	gcg tat Ala Tyr	4393
tgc ctg aca Cys Leu Thr 865		r Val Val					4441
gga aag ccg Gly Lys Pro							4489
gat gag atg Asp Glu Met		s Ala Ser					4537

atg cag cto Met Gln Leo 918	ı Ala Glu	Gln Phe					
aca gcc acc Thr Ala Thr 930							
tgg cgg acc Trp Arg Thi 945							
agc ggg ata Ser Gly Ile		_			_		Pro
gcg ata gca Ala Ile Ala							
acc acc caa Thr Thr Gli 99	His Thr	Leu Leu				Trp Val	
gcc caa ctt Ala Gln Let 1010			Ala Ala				
atc gct gga Ile Ala Gly 1025					Gly Lys		
gat att ttg Asp Ile Lew		Tyr Gly					Ala
ttt aag gto Phe Lys Val			_	Ser Thr			
cta ctc cct Leu Leu Pro 107	Ala Ile	Leu Ser				Gly Val	
tgc gca gcc Cys Ala Ala 1090							
cag tgg atg Gln Trp Met 1105					Arg Gly		
tcc ccc acc	cac tat	qtq cct (	gag agc	gac gct	gca gca	cgt gtc	act 5209

cag atc ctc tct agt Gln Ile Leu Ser Ser 1140			g Leu His
cag tgg atc aac gag Gln Trp Ile Asn Glu 1155	gac tgc tcc acg Asp Cys Ser Thr 1160	cca tgc tcc ggc tc Pro Cys Ser Gly Sei 1165	g tgg cta 5305 r Trp Leu
aga gat gtt tgg gat Arg Asp Val Trp Asp 1170			
tgg ctc cag tcc aag Trp Leu Gln Ser Lys 1185	ctc ctg ccg cga Leu Leu Pro Arg 1190	ttg ccg gga gtc ccc Leu Pro Gly Val Pro 1195	ttc ttc 5401 o Phe Phe 1200
tca tgt caa cgt ggg Ser Cys Gln Arg Gly 120	Tyr Lys Gly Val		
caa acc acc tgc cca Gln Thr Thr Cys Pro 1220			l Lys Asn
tgt tcc atg agg atc Cys Ser Met Arg Ile 1235	gtg ggg cct agg Val Gly Pro Arg 1240	acc tgt agt aac ac Thr Cys Ser Asn Th 1245	g tgg cat 5545 r Trp His
gga aca ttc ccc att Gly Thr Phe Pro Ile 1250			
ccg gcg cca aat tat Pro Ala Pro Asn Ty 1265			
tac gtg gag gtt acg Tyr Val Glu Val Thr 128	Arg Val Gly Asp		
acc act gac aac gta Thr Thr Asp Asn Val 1300			o Glu Phe
ttc aca gaa gtg gat Phe Thr Glu Val Asp 1315			
aaa ccc ctc cta cgg Lys Pro Leu Leu Arg 1330			
tac ctg gtt ggg tca Tyr Leu Val Gly Ser			

	ctc Leu			_	Leu		-			His		-			Thr	5929
	aag Lys			Leu					Pro					Ser	tca Ser	5977
	gct Ala		Gln	_				Ser	_	_	-		Cys			6025
	cat His 1410	Āsp					Āsp					Asn				6073
	cag Gln					Asn					Glu					6121
	gta Val				Ser					Gln					Glu	6169
	gaa Glu			Val					Leu					ГЛЗ		6217
	cga Arg		Met					Arg					Pro			6265
	gag Glu 1490	Ser					Asp					Val				6313
_	cca Pro	_	-		-	Гуs	_		_		Pro					6361
	agg Arg				Leu					Val					Ala	6409
	ctc Leu			Lys					Ser					Val		6457
	gly		Ala					Āsp					Asp			6505
	gga Gly 1570	Ser					Tyr					Pro				6553

	Pro					Leu			gly aaa		${\tt Trp}$					6601
					Asp				tgc Cys 1610	Ser					Trp	6649
				Ile					gcg Ala					Leu		6697
		_	Leu	_			-	Leu	cgt Arg				Leu	-		6745
_		Thr		-	_	_	Ser		cgg Arg	_	_	Lys	_			6793
Asp 1665	Arg	Leu	Gln	Val	Leu 1670	Asp )	Asp	His	tac Tyr	Arg 1675	Asp	Val	Leu	Lys	Glu 1680	6841
atg Met					Ser				gct Ala 1690	Lys					Glu	6889
_	-	_	_	Leu	_				tcg Ser	_	-			Phe		6937
		-	Lys	_	_			Leu	tcc Ser	_	-	_	Val			6985
		Ser			_		Leu	_	gaa Glu	_		Glu				7033
gac Asp 1745	Thr			_	-	Lys			gtt Val		Cys	_				7081
_			_		Pro	-	-		atc Ile 1770	Val			-		Gly	7129
gtt Val				Glu					Tyr					Thr		7177
cct Pro	_	_	Val	_				Tyr	gga Gly				Ser			7225

cag cgg gtc gag Gln Arg Val Glu 1810		aat gct tgg aaa Asn Ala Trp Lys		
atg ggc ttc gca Met Gly Phe Ala 1825	tat gac acc o Tyr Asp Thr 1	egc tgt ttt gac Arg Cys Phe Asp 1835	Ser Thr Val	act gag 7321 Thr Glu 1840
aat gac atc cgt Asn Asp Ile Arg			Cys Cys Asp	
ccc gaa gcc aga Pro Glu Ala Arg 186	Gln Ala Ile A			Tyr Ile
ggg ggc ccc ctg Gly Gly Pro Let 1875	Thr Asn Ser I			
		acg acc agc tgc Thr Thr Ser Cys		
tgt tac ttg aag Cys Tyr Leu Lys 1905		gee tgt ega get Ala Cys Arg Ala 1915	Ala Lys Leu	
		gac gac ctt gtc Asp Asp Leu Val 1930	Val Ile Cys	
gcg ggg acc caa Ala Gly Thr Glr 194	Glu Asp Glu A			Glu Ala
atg act aga tac Met Thr Arg Tyn 1955	Ser Ala Pro I	ect ggg gac ccg Pro Gly Asp Pro 1960		-
gac ttg gag ttg Asp Leu Glu Leu 1970		ggg too too aat Cys Ser Ser Asn		
		ac tat ctc acc Tyr Tyr Leu Thr 1995	Arg Asp Pro	
Asp Ala Ser Gly 1985 ccc ctt gcg cgg	Lys Arg Val 1 1990 get geg tgg g	Tyr Tyr Leu Thr	Arg Asp Pro cac act cca His Thr Pro	Thr Thr 2000 gtc aat 7849

atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa caa 7945  Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln 2035 2040 2045
ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc att 7993 Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile 2050 2055 2060
gag cca ctt gac cta cct cag atc att caa cga ctc cac ggc ctt agc 8041 Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser 2065 2070 2075 2080
gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg gct 8089 Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala 2085 2090 2095
tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat 8137 Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His 2100 2105 2110
cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct 8185 Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala 2115 2120 2125
gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc 8233 Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu 2130 2135 2140
aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg 8281 Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp 2145 2150 2155 2160
ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt 8329 Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg 2165 2170 2175
gcc cga ccc cgc tgg ttc acg tgg tgc cta ctc cta ctt tct gta ggg 8377 Ala Arg Pro Arg Trp Phe Thr Trp Cys Leu Leu Leu Leu Ser Val Gly 2180 2185 2190
gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca 8427 Val Gly Ile Tyr Leu Leu Pro Asn Arg * 2195 2200
ggccaatagg ccatcetgtt tttttccctt ttttcccttt tttttttt tttttt
<210> 6 <211> 8638 <212> DNA <213> HCV <220> <221> CDS <222> (1802)(8407)
< <del>4</del> 00> 6

accagecece	asttaaa.	ת מסמסמי	-dasa	nationat	-020	taccata	ta=	สสรรา	rtacto	60
tcttcacgca										
ccccacgea		_			_		_			
gacgaccggg										
gcgagactgc										
gtgcttgcga										
ctcaaagaaa										
cggccgcttg										
ctgatgccgc										
acctgtccgc										
cgacgggcgt							_			
tgctattggg							-			
aagtatccat										
cattcgacca		_					_	_		
ttgtcgatca										
ccaggeteaa										
gcttgccgaa										
tgggtgtggc										
ttggcggcga										
agcgcatcgc										
cagaccacaa										
ccgaagccgc										
geegtetttt										
taggggtctt										
agttcctctg										
gaacccccca										
tgcaaaggcg					_					
atgactetee	tcaagcgta		aaaaa				മനന	racce	:Carro	
atggetetee									_	
tatgggatct	gatctgggg	c ctcggl	gèac	atgettt	aca	tgtgttt	agt	cgago	gttaaa	1740
tatgggatet aaacgtetag	gatetgggg geeceeega	c ctcggi a ccacg	gèac gggac	atgcttt gtggttt	aca tcc	tgtgttt tttgaaa	agt aac	cgago acgat	gttaaa caatac	1740 1800
tatgggatet aaacgtetag c atg gae c	gatetgggg geeceega gg gag at	c ctcggt a ccacgg g gca go	gèac gggac ca tcg	atgettt gtggttt g tge gg	taca tcc ga gg	tgtgttt tttgaaa c gcg g	agt aac tt't	cgagg acgat tc gt	gttaaa caatac ca ggt	1740 1800
tatgggatet aaacgtetag	gatetgggg geeceega gg gag at	c ctcggt a ccacgg g gca go	gèac gggac ca tcg	atgettt gtggttt g tge gg c Cys Gl	taca tcc ga gg	tgtgttt tttgaaa c gcg g	agt aac tt't	cgagg acgat tc gt he Va	gttaaa caatac ca ggt	1740 1800
tatgggatet aaacgtetag c atg gac c Met Asp A	gatctgggg gcccccga gg gag at rg Glu Me	c ctcggt a ccacgg g gca go	gèac gggac ca tcg	atgettt gtggttt g tge gg c Cys Gl	aca tcc ga gg ly Gl	tgtgttt tttgaaa c gcg g	agt aac tt't	cgagg acgat tc gt he Va	gttaaa caatac ca ggt al Gly	1740 1800
tatgggatct aaacgtctag c atg gac c Met Asp A 1	gatetgggg geeeeega gg gag at rg Glu Me 5	c ctcggi a ccacgg g gca gc t Ala Al	gèac ggac ca tcg La Ser	atgettt gtggttt g tge gg Cys Gl	aca tcc ga gg ly Gl	tgtgttt tttgaaa c geg g y Ala V	agt aac tt't al P	cgagg acgat tc gt he Va	gttaaa aatac a ggt al Gly L5	1740 1800
tatgggatet aaacgtetag c atg gac c Met Asp A	gatctgggg gcccccga gg gag at rg Glu Me 5 ttg acc	c ctcggi a ccacgg g gca go t Ala Al ttg tca	gèac ggac ca teg la Ser	atgettt gtggttt g tge gg Cys Gl	aca tcc ga gg ly Gl LO aag	tgtgttt tttgaaa c gcg g y Ala V	agt aac tt't al P	cgagg acgat tc gt he Va gct	gttaaa caatac ca ggt al Gly L5	1740 1800 1849
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc	gatctgggg gcccccga gg gag at rg Glu Me 5 ttg acc	c ctcggi a ccacgg g gca go t Ala Al ttg tca	gèac ggac ca tog la Ser	atgettt gtggttt g tge gg Cys Gl	aca tcc ga gg ly Gl LO aag	tgtgttt tttgaaa c gcg g y Ala V	agt aac tt't al P	cgagg acgat tc gt he Va gct	gttaaa caatac ca ggt al Gly L5	1740 1800 1849
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc	gatctgggg gccccccga gg gag at rg Glu Me 5 ttg acc Leu Thr	c ctcggi a ccacgg g gca go t Ala Al ttg tca	gèac ggac ca tog la Ser	atgettt gtggttt g tge gg Cys Gl cac tat lis Tyr	aca tcc ga gg ly Gl LO aag	tgtgttt tttgaaa c gcg g y Ala V	agt aac tt't al P ctc Leu	cgagg acgat tc gt he Va gct	gttaaa caatac ca ggt al Gly L5	1740 1800 1849
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc	gatctgggg gccccccga gg gag at rg Glu Me 5 ttg acc Leu Thr	c cteggi a ccaegg g gca go t Ala Al ttg tca Leu Ser	ggac ggac ca tcg la Ser ccg c	atgettt gtggttt g tge gg c Cys GJ cac tat His Tyr 25	aca tcc ja gg ly Gl LO aag Lys	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe	agt aac tt't al P ctc Leu 30	cgagg acgat tc gt he Va gct Ala	gttaaa caatac ca ggt al Gly L5 agg Arg	1740 1800 1849
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc Leu Ile Leu	gatctgggg gcccccga gg gag at rg Glu Me 5 ttg acc Leu Thr 20	c ctcggl a ccacgg g gca gc t Ala Al ttg tca Leu Ser	ogdac gggac ca tog la Ser cog c Pro H	atgettt gtggttt g tge gg c Cys Gl cac tat His Tyr 25	aca tec ya gg ly Gl LO aag Lys	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe	agt aac tt't al P ctc Leu 30	cgagg acgat tc gt he Va gct Ala	gttaaa caatac ca ggt al Gly L5 agg Arg	1740 1800 1849 1897
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc Leu Ile Leu ctc ata tgg	gatctgggg gcccccga gg gag at rg Glu Me 5 ttg acc Leu Thr 20 tgg tta	c ctcggl a ccacgg g gca gc t Ala Al ttg tca Leu Ser	ogdac gggac ca tog la Ser cog c Pro H	atgettt gtggttt g tge gg c Cys Gl cac tat His Tyr 25	aca tec ya gg ly Gl LO aag Lys	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe	agt aac tt't al P ctc Leu 30	cgagg acgat tc gt he Va gct Ala	gttaaa caatac ca ggt al Gly L5 agg Arg	1740 1800 1849 1897
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc Leu Ile Leu ctc ata tgg Leu Ile Trp	gatctgggg gcccccga gg gag at rg Glu Me 5 ttg acc Leu Thr 20 tgg tta	c ctcggl a ccacgg g gca gc t Ala Al ttg tca Leu Ser	cgcac ggac ca tcg la Ser cog c Pro H ttt a Phe I	atgettt gtggttt g tge gg c Cys Gl cac tat His Tyr 25	aca tec ya gg ly Gl LO aag Lys	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu	agt aac tt't al P ctc Leu 30	cgagg acgat tc gt he Va gct Ala	gttaaa caatac ca ggt al Gly L5 agg Arg	1740 1800 1849 1897
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc Leu Ile Leu ctc ata tgg Leu Ile Trp	gatctgggggggcccccga gg gag at rg Glu Me 5 ttg acc Leu Thr: 20 tgg tta	c ctcggla ccacgg g gca gc t Ala Al  ttg tca Leu Ser  caa tat Gln Tyr	cgcac gggac ca tcg la Ser cog c Pro H ttt a Phe I 40	atgettt gtggttt g tge gg c Cys Gl cac tat His Tyr 25 atc acc	aca tcc ya gg ly Gl LO aag Lys agg Arg	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45	agt aac tt't al P ctc Leu 30 gca Ala	cgagg acgat tc gt he Va gct Ala cac	gttaaa caatac ca ggt al Gly L5 agg Arg ttg Leu	1740 1800 1849 1897
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc Leu Ile Leu ctc ata tgg Leu Ile Trp 35	gatctgggggggcccccga gg gag at rg Glu Me 5 ttg acc Leu Thr 20 tgg tta Trp Leu	c cteggia ccacgo g gca go t Ala Al  ttg tca Leu Ser  caa tat Gln Tyr	cgcac ggac ca tcg la Ser cog c Pro H ttt a Phe I 40 aac g	atgettt gtggttt g tge gg c Cys Gl cac tat His Tyr 25 atc acc He Thr	aca tcc ya gg ly G1 L0 aag Lys agg Arg	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45 ggc cgc	agt agt agc at tt tal P ctc Leu 30 gca Ala	cgagg acgat tc gt he Va gct Ala cac His	gttaaa caatac ca ggt al Gly L5 agg Arg ttg Leu	1740 1800 1849 1897
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc Leu Ile Leu ctc ata tgg Leu Ile Trp 35 caa gtg tgg	gatctgggggggcccccga gg gag at rg Glu Me 5 ttg acc Leu Thr 20 tgg tta Trp Leu	c cteggia ccacgo g gca go t Ala Al  ttg tca Leu Ser  caa tat Gln Tyr	cgcac ggac ca tcg la Ser cog c Pro H ttt a Phe I 40 aac g	atgettt gtggttt g tge gg c Cys Gl cac tat His Tyr 25 atc acc He Thr	aca tcc ya gg ly G1 L0 aag Lys agg Arg	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45 ggc cgc	agt agt agc at tt tal P ctc Leu 30 gca Ala	cgagg acgat tc gt he Va gct Ala cac His	gttaaa caatac ca ggt al Gly L5 agg Arg ttg Leu	1740 1800 1849 1897
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc Leu Ile Leu ctc ata tgg Leu Ile Trp 35 caa gtg tgg Gln Val Trp	gatctgggggggcccccga gg gag at rg Glu Me 5 ttg acc Leu Thr 20 tgg tta Trp Leu	c ctcggla ccacgg g gca gc t Ala Al  ttg tca Leu Ser  caa tat Gln Tyr  ccc ctc Pro Leu	cgcac ggac ca tcg la Ser cog c Pro H ttt a Phe I 40 aac g	atgettt gtggttt g tge gg c Cys Gl cac tat His Tyr 25 atc acc He Thr	aca tcc ya gg ly G1 L0 aag Lys agg Arg	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45 ggc cgc Gly Arg	agt agt agc at tt tal P ctc Leu 30 gca Ala	cgagg acgat tc gt he Va gct Ala cac His	gttaaa caatac ca ggt al Gly L5 agg Arg ttg Leu	1740 1800 1849 1897
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc Leu Ile Leu ctc ata tgg Leu Ile Trp 35 caa gtg tgg Gln Val Trp	gatctggggggcccccga gg gag at rg Glu Me 5 ttg acc Leu Thr 20 tgg tta Trp Leu atc ccc Ile Pro	c ctcggla ccacgg g gca gc t Ala Al  ttg tca Leu Ser  caa tat Gln Tyr  ccc ctc Pro Leu 55	cgcac ggac ca tcg la Ser ccg c Pro H ttt a Phe I 40 aac g Asn V	atgettt gtggttt gtggttt gtgc gg Cys Gl cac tat His Tyr 25 Atc acc Le Thr gtt cgg Arg	aca tcc fa gg ly Gl Lys agg Arg Arg	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45 ggc cgc Gly Arg 60	agt agt agc tt't al P ctc Leu 30 gca Ala gat	cgagg acgat tc gt he Va gct Ala cac His	gttaaa caatac ca ggt al Gly L5 agg Arg ttg Leu gtc Val	1740 1800 1849 1897
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc Leu Ile Leu ctc ata tgg Leu Ile Trp 35 caa gtg tgg Gln Val Trp 50	gatctgggggggcccccga gg gag at rg Glu Me 5 ttg acc Leu Thr 20 tgg tta Trp Leu atc ccc Ile Pro	c cteggia ccacgo g gca gc t Ala Al  ttg tca Leu Ser  caa tat Gln Tyr  ccc ctc Pro Leu 55 gcg atc	cgcac ggac ca tcg la Ser cog c Pro H ttt a Phe I 40 aac g Asn V	atgettt gtggttt gtggttt gtgc gg Cys Gl cac tat His Tyr 25 atc acc He Thr gtt cgg Arg	aca tcc ga gg ly Gl LO aag Lys agg Arg Gly	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45 ggc cgc Gly Arg 60 atc ttt	agt agt agc agt ctc ti't al P ctc Leu 30 gca Ala gat Asp	cgagg acgat tc gt he Va gct Ala cac His gcc Ala	gttaaa caatac ca ggt al Gly L5 agg Arg ttg Leu gtc Val	1740 1800 1849 1897 1945
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc Leu Ile Leu ctc ata tgg Leu Ile Trp 35 caa gtg tgg Gln Val Trp 50 atc ctc ctc	gatctgggggggcccccga gg gag at rg Glu Me 5 ttg acc Leu Thr 20 tgg tta Trp Leu atc ccc Ile Pro	c cteggia ccacgo g gca gc t Ala Al  ttg tca Leu Ser  caa tat Gln Tyr  ccc ctc Pro Leu 55 gcg atc	cgcac ggac ca tcg la Ser cog c Pro H ttt a Phe I 40 aac g Asn V	atgettt gtggttt gtggttt gtgc gg Cys Gl cac tat His Tyr 25 atc acc He Thr gtt cgg Arg	aca tcc ga gg ly Gl LO aag Lys agg Arg Gly	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45 ggc cgc Gly Arg 60 atc ttt	agt agt agc agt ctc ti't al P ctc Leu 30 gca Ala gat Asp	cgagg acgat tc gt he Va gct Ala cac His gcc Ala	gttaaa caatac ca ggt al Gly L5 agg Arg ttg Leu gtc Val	1740 1800 1849 1897 1945
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc Leu Ile Leu ctc ata tgg Leu Ile Trp 35 caa gtg tgg Gln Val Trp 50 atc ctc ctc Ile Leu Leu	gatctgggggggcccccga gg gag at rg Glu Me 5 ttg acc Leu Thr 20 tgg tta Trp Leu atc ccc Ile Pro	c ctcggda ccacggg gca gca tat Ser Caa tat Ser Ccc ctc Pro Leu 55 gcg atc Ala Ile	cgcac ggac ca tcg la Ser cog c Pro H ttt a Phe I 40 aac g Asn V	atgettt gtggttt gtggttt gtgc gg Cys Gl cac tat His Tyr 25 atc acc He Thr gtt cgg Arg	aca tcc ga gg ly Gl LO aag Lys agg Arg Gly cta Leu	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45 ggc cgc Gly Arg 60 atc ttt	agt agt agc agt ctc ti't al P ctc Leu 30 gca Ala gat Asp	cgagg acgat tc gt he Va gct Ala cac His gcc Ala	gttaaa caatac ca ggt al Gly L5 agg Arg ttg Leu gtc Val	1740 1800 1849 1897 1945
tatgggatct aaacgtctag c atg gac c Met Asp A 1  ctg ata ctc Leu Ile Leu  ctc ata tgg Leu Ile Trp 35  caa gtg tgg Gln Val Trp 50  atc ctc ctc Ile Leu Leu 65	gatctgggg gcccccga gg gag at rg Glu Me 5  ttg acc Leu Thr: 20  tgg tta Trp Leu atc ccc Ile Pro: acg tgc	c ctcggda ccacgg gca gca gca tat Ser cca tat Sln Tyr ccc ctc Pro Leu 55 gcg atc Ala Ile 70	cgcac ggac ca tcg la Ser ccg c Pro H ttt a Phe I 40 aac g Asn V cac c His P	atgettt gtggttt gtggttt gtgc gg Cys Gl cac tat lis Tyr 25 atc acc le Thr gtt cgg Val Arg	aca tec ga gg ly Gl LO aag Lys agg Arg Gly cta Leu 75	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45 ggc cgc Gly Arg 60 atc ttt Ile Phe	agt aac tt t al P ctc Leu 30 gca Ala gat Asp	cgagg acgat tc gt he Va gct Ala cac His gcc Ala	gttaaa aatac a ggt al Gly L5 agg Arg ttg Leu gtc Val acc Thr 80	1740 1800 1849 1897 1945
tatgggatct aaacgtctag c atg gac c Met Asp A 1 ctg ata ctc Leu Ile Leu ctc ata tgg Leu Ile Trp 35 caa gtg tgg Gln Val Trp 50 atc ctc ctc Ile Leu Leu	gatctggggggeccccga gg gag at rg Glu Me 5  ttg acc Leu Thr: 20  tgg tta Trp Leu atc ccc Ile Pro: acg tgc; Thr Cys;	c ctcggga ccacgg gca gca gca tat Ser caa tat Sin Tyr ccc ctc Pro Leu 55 gcg atc Ala Ile 70 ata ctc	cgcac ggac ca tcg la Ser ccg c Pro H ttt a Phe I 40 aac g Asn V cac c His P	atgettt gtggttt gtggttt gtgc gg Cys GJ cac tat His Tyr 25 Atc acc He Thr gtt cgg Cal Arg Ca gag Cro Glu cca ctc	aca tcc ga gg ly Gl l0 aag Lys agg Arg ggg Gly cta Leu 75 atg	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45 ggc cgc Gly Arg 60 atc ttt Ile Phe	agt agc	cgaggatacgattc gthe Value Ala cac His gcc Ala atc Ile	gttaaa aatac a ggt al Gly L5 agg Arg ttg Leu gtc Val acc Thr 80 ggt	1740 1800 1849 1897 1945 1993
tatgggatct aaacgtctag c atg gac c Met Asp A 1  ctg ata ctc Leu Ile Leu  ctc ata tgg Leu Ile Trp 35  caa gtg tgg Gln Val Trp 50  atc ctc ctc Ile Leu Leu 65  aaa atc ttg	gatctggggggeccccga gg gag at rg Glu Me 5  ttg acc Leu Thr: 20  tgg tta Trp Leu atc ccc Ile Pro: acg tgc; Thr Cys;	c ctcggga ccacgg gca gca gca tat Ser caa tat Sin Tyr ccc ctc Pro Leu 55 gcg atc Ala Ile 70 ata ctc	cgcac ggac ca tcg la Ser ccg c Pro H ttt a Phe I 40 aac g Asn V cac c His P	atgettt gtggttt gtggttt gtgc gg Cys GJ cac tat His Tyr 25 Atc acc He Thr gtt cgg Cal Arg Ca gag Cro Glu cca ctc	aca tcc ga gg ly G1 l0 aag Lys agg Arg ggg Gly cta Leu 75 atg	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45 ggc cgc Gly Arg 60 atc ttt Ile Phe	agt agc	cgaggatacgattc gthe Value Ala cac His gcc Ala atc Ile	gttaaa aatac a ggt al Gly L5 agg Arg ttg Leu gtc Val acc Thr 80 ggt	1740 1800 1849 1897 1945 1993
tatgggatct aaacgtctag c atg gac c Met Asp A 1  ctg ata ctc Leu Ile Leu  ctc ata tgg Leu Ile Trp 35  caa gtg tgg Gln Val Trp 50  atc ctc ctc Ile Leu Leu 65  aaa atc ttg	gatctgggg gcccccga gg gag at rg Glu Me 5  ttg acc Leu Thr: 20  tgg tta Trp Leu atc ccc Ile Pro: acg tgc; Thr Cys; ctc gcc; Leu Ala	c ctcggga ccacgg gca gca gca tat Ser caa tat Sin Tyr ccc ctc Pro Leu 55 gcg atc Ala Ile 70 ata ctc	cgcac ggac ca tcg la Ser ccg c Pro H ttt a Phe I 40 aac g Asn V cac c His P	atgettt gtggttt gtggttt gtgc gg Cys Gl cac tat lis Tyr 25 atc acc le Thr gtt cgg Cal Arg ca gag Cro Glu cca ctc Cro Leu	aca tcc ga gg ly G1 l0 aag Lys agg Arg ggg Gly cta Leu 75 atg	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45 ggc cgc Gly Arg 60 atc ttt Ile Phe	agt agc	cgaggatacgattc gthe Value Gac Ala atc Ile gct Ala	gttaaa aatac a ggt al Gly L5 agg Arg ttg Leu gtc Val acc Thr 80 ggt	1740 1800 1849 1897 1945 1993
tatgggatct aaacgtctag c atg gac c Met Asp A 1  ctg ata ctc Leu Ile Leu  ctc ata tgg Leu Ile Trp 35  caa gtg tgg Gln Val Trp 50  atc ctc ctc Ile Leu Leu 65  aaa atc ttg Lys Ile Leu	gatctgggg gcccccga gg gag at rg Glu Me 5  ttg acc Leu Thr 20  tgg tta Trp Leu atc ccc Ile Pro acg tgc Thr Cys ctc gcc Leu Ala 85	c ctcggda ccacggg gca gca gca tat stem Ser caa tat sln Tyr ccc ctc Pro Leu 55 gcg atc Ala Ile 70 ata ctc Ile Leu	cgcac ggac ca tcg la Ser ccg c Pro H ttt a Phe I 40 aac g Asn V cac c His P	atgettt gtggttt gtggttt gtgc gg Cys GJ  cac tat His Tyr 25 Atc acc Thr  gtt cgg Val Arg  cca gag Cro Glu  cca ctc Pro Leu 90	aca tcc ga gg ly Gl l0 aag Lys agg Arg Gly cta Leu 75 atg Met	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45 ggc cgc Gly Arg 60 atc ttt Ile Phe gtg ctc Val Leu	agt agc agc agc tt'tal P ctc Leu 30 gca Ala gat Asp acc Thr cag	cgaggatacgattc gthe Value Cac His gcc Ala atc Ile gct Ala 95	gttaaa aatac a ggt al Gly L5 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly	1740 1800 1849 1897 1945 1993
tatgggatct aaacgtctag c atg gac c Met Asp A 1  ctg ata ctc Leu Ile Leu  ctc ata tgg Leu Ile Trp 35  caa gtg tgg Gln Val Trp 50  atc ctc ctc Ile Leu Leu 65  aaa atc ttg Lys Ile Leu  ata acc aaa	gatctgggg gcccccga gg gag at rg Glu Me 5  ttg acc Leu Thr 20  tgg tta Trp Leu atc ccc Ile Pro acg tgc Thr Cys ctc gcc Leu Ala 85	c ctcggda ccacggg gca gca gca tat seu Ser caa tat sin Tyr ccc ctc Pro Leu 55 gcg atc Ala Ile 70 ata ctc Ile Leu tac ttc	cgdac gggac ca tcg la Ser ccg c Pro H ttt a Phe I 40 aac g Asn V cac c His P	atgettt gtggttt gtggttt gtgc gg Cys Gl Cac tat His Tyr 25 Atc acc Thr Atc acc Thr Gtt cgg Coa gag Cro Glu Coa ctc Cro Leu 90 Cgc gca	aca tcc ga gg ly Gl l0 aag Lys agg Arg ggg Gly cta Leu 75 atg Met	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45 ggc cgc Gly Arg 60 atc ttt Ile Phe gtg ctc Val Leu	agt agc agc tt'tal P ctc Leu 30 gca Ala gat Asp acc Thr cag Gln att	cgaggatacgattc gthe Value Cac Ala atc Ile gct Ala 95	gttaaa aatac a ggt al Gly L5 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly	1740 1800 1849 1897 1945 1993 2041
tatgggatct aaacgtctag c atg gac c Met Asp A 1  ctg ata ctc Leu Ile Leu  ctc ata tgg Leu Ile Trp 35  caa gtg tgg Gln Val Trp 50  atc ctc ctc Ile Leu Leu 65  aaa atc ttg Lys Ile Leu	gatctgggg gcccccga gg gag at rg Glu Me 5  ttg acc Leu Thr 20  tgg tta Trp Leu atc ccc Ile Pro acg tgc Thr Cys ctc gcc Leu Ala 85	c ctcggda ccacggg gca gca gca tat seu Ser caa tat sin Tyr ccc ctc Pro Leu 55 gcg atc Ala Ile 70 ata ctc Ile Leu tac ttc	cgcac ggac ca tcg la Ser ccg c Pro H ttt a Phe I 40 aac g Asn V cac c His P ggt c Gly P	atgettt gtggttt gtggttt gtgc gg Cys Gl Cac tat His Tyr 25 Atc acc Thr Atc acc Thr Gtt cgg Coa gag Cro Glu Coa ctc Cro Leu 90 Cgc gca	aca tcc ga gg ly Gl l0 aag Lys agg Arg ggg Gly cta Leu 75 atg Met	tgtgttt tttgaaa c gcg g y Ala V ctg ttc Leu Phe gcc gag Ala Glu 45 ggc cgc Gly Arg 60 atc ttt Ile Phe gtg ctc Val Leu	agt agc agc tt'tal P ctc Leu 30 gca Ala gat Asp acc Thr cag Gln att	cgaggatacgattc gthe Value Cac Ala atc Ile gct Ala 95	gttaaa aatac a ggt al Gly L5 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly	1740 1800 1849 1897 1945 1993 2041

												gtc Val 125				2185
												tat Tyr				2233
												gac Asp				2281
gca Ala	gtt Val	gag Glu	ccc Pro	gtc Val 165	gtc Val	ttc Phe	tct Ser	gat Asp	atg Met 170	gag Glu	acc Thr	aag Lys	gtt Val	atc Ile 175	acc Thr	2329
_			-						-			ttg Leu		_		2377
												ccg Pro 205				2425
	-		_			_						acg Thr				2473
	_	_	_					_				agc Ser				2521
												gtc Val				2569
												tgt Cys				2617
tat Tyr	cat His	ggt Gly 275	gcc Ala	ggc ggc	tca Ser	aag Lys	acc Thr 280	ctt Leu	gcc Ala	ggc	cca Pro	aag Lys 285	Gly	cca Pro	atc Ile	2665
_	_			_			-		_		-	gly	~~			2713
												ggc				2761
ctt	+	tta	atc	aco	agg	cat	gcc	gat	atc	att	cca	ata	cac	caa	raa	2809

ggc ggc	gac Asp	gly ggc	agg Arg 340	gjå aaa	agc Ser	cta Leu	ctc Leu	tcc Ser 345	ccc Pro	agg Arg	ccc Pro	gtc Val	tcc Ser 350	tac Tyr	ttg Leu	2857
						cca Pro										2905
						gtg Val 375										2953
						tct Ser										3001
						cct Pro										3049
_						act Thr		_		-						3097
						gly ggg										3145
						ttc Phe 455										3193
atc Ile 465	gac Asp	cct Pro	aac Asn	atc Ile	aga Arg 470	acc Thr	gly aaa	gta Val	agg Arg	acc Thr 475	atc Ile	acc Thr	acg Thr	ggt Gly	gcc Ala 480	3241
		_				tat Tyr		_			_					3289
						atc Ile										3337
gac Asp	tcg Ser	acc Thr 515	act Thr	atc Ile	ctg Leu	ggc	atc Ile 520	ggc	aca Thr	gtc Val	ctg Leu	gac Asp 525	caa Gln	gcg Ala	gag Glu	3385
						gtc Val 535										3433
						cca Pro									agc Ser 560	3481

										atc Ile						3529
aag Lys	eja aaa	gjå aaa	agg Arg 580	cac His	ctc Leu	att Ile	ttc Phe	tgc Cys 585	cat His	tcc Ser	aag Lys	aag Lys	aaa Lys 590	tgt Cys	gat Asp	3577
gag Glu	ctc Leu	gcc Ala 595	gcg Ala	aag Lys	ctg Leu	tcc Ser	ggc 600	ctc Leu	gga Gly	ctc Leu	aat Asn	gct Ala 605	gta Val	gca Ala	tat Tyr	3625
, tac Tyr										act Thr						3673
										ttt Phe 635						3721
tca Ser	gtg Val	atc Ile	gac Asp	tgc Cys 645	aat Asn	aca Thr	tgt Cys	gtc Val	acc Thr 650	cag Gln	aca Thr	gtc Val	gac Asp	ttc Phe 655	agc Ser	3769
										acc Thr						3817
										ggt Gly						3865
										ccc Pro						3913
										ggc Gly 715						3961
			_	~ ~			_			cgg Arg	_					4009
cca		++c	~~~		taa	cad	qac	cat	ctg	gag	ttc	tgg	gag	agc	gtc	4057
		_	_				_		Leu	Glu		Trp	Glu 750	Ser	Val	
Pro ttt	Gly	Leu	Pro 740 ctc	Val	Cys	Gln	Asp	His 745 gcc	cat		Phe ttg	tac	750 cag	act	aag	4105

tgc gcc Cys Ala 785															4201
tgt cto				_		_	_				_		_	-	4249
tat agg	Leu		-	_				-							4297
acc aaa Thr Lya															4345
agc acc Ser Th: 85	Trp														4393
tgc ctg Cys Let 865				_		_							Leu		4441
gga aag Gly Ly	Pro	Ala	Ile 885	Ile	Pro	Asp	Arg	Glu 890	Val	Phe	Tyr	Arg	Glu 895	Phe	4489
gat gag Asp Gl	ı Met														4537
atg ca Met Gl		_	-												4585
aca gco Thr Ala 93	a Thr														4633
tgg cgg Trp Arg 945	-		-	_				_		_					4681
agc ggg Ser Gl								Ser							4729
	, 116		965					970					273		
gcg at Ala Il	a gca e Ala	tca	ctg					gaa					ccg		4777

gcc caa ct Ala Gln Le 1010	-		_	-			4873
atc gct gg Ile Ala Gl 1025	/ Ala Ala V						4921
gat att tte Asp Ile Le				Ala Gly			4969
ttt aag gt Phe Lys Va						Val Asn	5017
cta ctc cc Leu Leu Pro 10	Ala Ile L		Gly Ala	Leu Val			5065
tgc gca gc Cys Ala Ala 1090							5113
cag tgg atg Gln Trp Me 1105	Asn Arg L		-			•	5161
tee eee aeg Ser Pro Th				Ala Ala		-	5209
cag atc cto	-		_			Leu His	5257
cag tgg ato Gln Trp Ilo 11!	Asn Glu A	_	Thr Pro	Cys Ser			5305
aga gat gt Arg Asp Val							5353
tgg ctc cag Trp Leu Gli 1185	Ser Lys L				_		5401
tca tgt cas Ser Cys Gli				Arg Gly			5449
caa acc acc							5497

1235			tgt agt aac acg Cys Ser Asn Thr 1245	
		Tyr Thr Thr	ggc ccc tgc acc Gly Pro Cys Thr 1260	
			cgg gtg gct gct Arg Val Ala Ala 1275	
			cac tac gtg acg His Tyr Val Thr O	
_	Val Lys Cys		gtt ccg gcc ccc Val Pro Ala Pro 131	Glu Phe
			agg tac gct cca Arg Tyr Ala Pro 1325	
aaa ccc ctc cta Lys Pro Leu Leu 1330	cgg gag gag Arg Glu Glu 133	Val Thr Phe	ctg gtc ggg ctc Leu Val Gly Leu 1340	e aat caa 5833 Asn Gln
	_		ccc gaa ccg gac Pro Glu Pro Asp 1355	
ata ata aat taa	atg ctc acc			
	Met Leu Thr 1365	-	cac att acg gcg His Ile Thr Ala O	
Val Leu Thr Ser	1365 ctg gcc agg Leu Ala Arg	Asp Pro Ser 1370 gga tct ccc	His Ile Thr Ala	Glu Thr 1375 agc tca 5977 Ser Ser
Val Leu Thr Ser gct aag cgt agg Ala Lys Arg Arg 138 tca gct agc cag	1365 ctg gcc agg Leu Ala Arg 0 ctg tct gcg	Asp Pro Ser 1370 gga tct ccc Gly Ser Pro 1385 cct tcc ttg	His Ile Thr Ala ccc tcc ttg gcc Pro Ser Leu Ala	e agc tca 5977 Ser Ser
yal Leu Thr Ser  gct aag cgt agg Ala Lys Arg Arg 138  tca gct agc cag Ser Ala Ser Glr 1395  cgt cat gac tcc	ctg gcc agg Leu Ala Arg  ctg tct gcg Leu Ser Ala	Asp Pro Ser 1370  gga tct ccc Gly Ser Pro 1385  cct tcc ttg Pro Ser Leu 1400  gac ctc atc Asp Leu Ile	His Ile Thr Ala  ccc tcc ttg gcc Pro Ser Leu Ala 139  aag gca aca tgc Lys Ala Thr Cys	e agc tca 5977 Ser Ser ser to 6025 Thr Thr
gct aag cgt agg Ala Lys Arg Arg 138 tca gct agc cag Ser Ala Ser Glr 1395 cgt cat gac tcc Arg His Asp Ser 1410 cgg cag gag atg	ctg gcc agg Leu Ala Arg  ctg tct gcg Leu Ser Ala  ccg gac gct Pro Asp Ala 141:	Asp Pro Ser 1370  gga tct ccc Gly Ser Pro 1385  cct tcc ttg Pro Ser Leu 1400  gac ctc atc Asp Leu Ile 5	Ccc tcc ttg gcc Pro Ser Leu Ala 139 aag gca aca tgc Lys Ala Thr Cys 1405 gag gcc aac ctc Glu Ala Asn Leu	e agc tca 5977 Ser Ser 90 e act acc 6025 Thr Thr ctg tgg 6073 Leu Trp aat aag 6121

agg gaa gta tcc g Arg Glu Val Ser Va 1460				Lys Phe
cct cga gcg atg co Pro Arg Ala Met Pr 1475		Arg Pro Asp		
tta gag tcc tgg as Leu Glu Ser Trp Ly 1490				
tgt cca ctg ccg cc Cys Pro Leu Pro Pr 1505		_	Pro Pro Pro	
aag agg acg gtt gt Lys Arg Thr Val Va 19				
gag ctc gcc aca as Glu Leu Ala Thr Ly 1540		-		Val Asp
agc ggc acg gca ac Ser Gly Thr Ala Ti 1555		Asp Gln Pro		
gcg gga tcc gac g Ala Gly Ser Asp Va 1570				
gag ccg ggg gat cc Glu Pro Gly Asp Pi 1585			Trp Ser Thr	
gag gag get agt ga Glu Glu Ala Ser G	g gac gtc gtc u Asp Val Val 05	tgc tgc tcg Cys Cys Ser 1610	atg tcc tac Met Ser Tyr	aca tgg 6649 Thr Trp 1615
aca ggc gcc ctg at Thr Gly Ala Leu I 1620				Leu Pro
atc aat gca ctg as Ile Asn Ala Leu Se 1635		Leu Arg His		
gct aca aca tct co Ala Thr Thr Ser An 1650	c agc gca agc g Ser Ala Ser 1655	: ctg cgg cag : Leu Arg Gln	aag aag gtc Lys Lys Val 1660	acc ttt 6793 Thr Phe
gac aga ctg cag gt Asp Arg Leu Gln Va 1665			Asp Val Leu	

	aag Lys				Ser					Lys					Glu	6889
	gcc Ala			Leu					Ser					Phe		6937
	G1y ggg		Lys					Leu					Val			6985
	cgc Arg 1730	Ser					Leu					Glu				7033
	acc Thr 5					Lys					Cys					7081
	Gly 999				Pro					Val					${ t Gly}$	7129
	cgt Arg			Glu					Tyr					Thr		7177
	cag Gln	_	Val	_				Tyr					Ser			7225
	cgg Arg 1810	Val					Asn					Lys				7273
-	ggc Gly 5		_		_	Thr		-	_	-	Ser		-			7321
	gac Asp				Glu					${\tt Gln}$					Ala	7369
	gaa Glu			Gln					Leu					Tyr		7417
	ggc		Leu					Gly					$\mathtt{Tyr}$		cgg Arg	7465
	cgc Arg 1890	Ala					Thr					Asn				7513

tgt tac ttg aag gc Cys Tyr Leu Lys Al 1905		
tgc acg atg ctc gt Cys Thr Met Leu Va 19	l Cys Gly Asp A	
geg ggg acc caa ga Ala Gly Thr Gln Gl 1940	ı Asp Glu Ala S	
atg act aga tac to Met Thr Arg Tyr Se 1955		
gac ttg gag ttg at Asp Leu Glu Leu Il 1970		 Ser Val Ala His
gat gca tct ggc aa Asp Ala Ser Gly Ly 1985		
ccc ctt gcg cgg gc Pro Leu Ala Arg Al 20	a Ala Trp Glu T	
tee tgg cta ggc aa Ser Trp Leu Gly As 2020	n Ile Ile Met T	
atg atc ctg atg ac Met Ile Leu Met Th 2035		
ctt gaa aaa gcc ct Leu Glu Lys Ala Le 2050		 Cys Tyr Ser Ile
gag cca ctt gac ct Glu Pro Leu Asp Le 2065		
gca ttt tca ctc ca Ala Phe Ser Leu Hi 20	s Ser Tyr Ser P	
tca tgc ctc agg aa Ser Cys Leu Arg Ly 2100	Leu Gly Val P	

gcc act tgt Ala Thr Cys 2130				Phe					Arg				8233
aaa ctc act Lys Leu Tho 2145			o Ala					Asp					8281
ttc gtt gct Phe Val Ala	a Gly T	ac ag yr Se 165	c Gly c ggg	gga Gly	gac Asp	ata Ile 2170	Tyr	cac His	agc Ser	ctg Leu	tct Ser 217	Arg	8329
gcc cga ccc Ala Arg Pro						Leu					Val		8377
gta ggc ato Val Gly Ile 219	e Tyr L				Arg	tġa *	acgg	gggag	get a	aaac	actco	ca	8427
ggccaatagg ttttttttt ccatcttagc gtgctgatac	tttttt cctagt	tttt cacg	ttttt! gctag	taata ctgtg	ttt gaaa	tttt	tcc	tttt	cctt	ccc	tttg	gtggct	8547
<210> 7 <211> 8638 <212> DNA <213> HCV													
<220> <221> CDS <222> (1802)	2)(8	407)											
<400> 7							•						
gccagccccc													
tcttcacgca													
ccccctccc gacgaccggg	yyyaya	geca	tag tg	30000	, cgg	jaacc	-gg-	gagı	-acat	ocy :	gaacı	racaaa	240
gcgagactgc													
gtgcttqcga													
ctcaaagaaa				_	-			_	-				
eggeegettg				_	_	-	_		_				
ctgatgccgc													
acctgtccgg													
cgacgggcgt	tccttg	cgca ;	gctgt	gctcg	g acc	jttgt	cac	tgaa	agcgg	ga a	aggga	actggc	660
tgctattggg	cgaagt	.gccg	gggcag	ggato	: tac	tgto	catc	tcac	cctto	gct :	cctgo	ccgaga	720
aagtatccat													
cattcgacca	ccaagc	gaaa '	catego	catcg	ago	gago	cacg	tact	cgga	atg 9	gaago	ceggte	840
ttgtcgatca													
ccaggeteaa gettgeegaa													
tgggtgtggc													
ttggcggcga													
agcgcatcgc													
cagaccacaa													
ccgaagccgc													
gccgtctttt													
5 5													

agt gaa tgc atg tat aaa c a	teet .cccc .aaag gete .ggga .cgte .tg g	etg geg tee tet tag ac c	gaag cctg gcac tcaa gatc gccc gg g	cttc gcga aacc gcgt tggg cccg ag a lu M	tt g ca g cc a at t gc c aa c tg g	aaga gtgc gtgc caac tcgg cacg ca g	caaa ctct cacg aagg tgca ggga ca t	c aa g cg t tg g gc c at c gt	cgtc gcca tgag tgaa gctt gctt gc g	tgta aaag ttgg ggat taca ttcc ga g	gcg cca ata gcc tgt ttt gc g	acco cgtg gttg caga gttt gaaa cg g	ttt tat tgg agg agt att t	gcag aaga aaag tacc cgag acga tc g	ggaagc gcagcg tacacc agtcaa ccattg gttaaa taatac ta ggt al Gly	1500 1560 1620 1680 1740 1800
										aag Lys				Ala		1897
										agg Arg						1945
caa Gln	gtg Val 50	Trp	atc Ile	ccc Pro	ccc Pro	ctc Leu 55	Asn	gtt Val	cgg Arg	gja aaa	ggc 60	cgc Arg	gat Asp	gcc Ala	gtc Val	1993
										cta Leu 75						2041
										atg Met						2089
										cac His						2137
					_	_	_			cat His		_		_	_	2185
										tac Tyr						2233
										cta Leu 155						2281
gca Ala	gtt Val	gag Glu	ccc Pro	gtc Val 165	gtc Val	ttc Phe	tct Ser	gat Asp	atg Met 170	gag Glu	acc Thr	aag Lys	gtt <b>Val</b>	atc Ile 175	acc Thr	2329
										atc Ile						2377
										ctg Leu						2425

ctt Leu	gaa Glu 210	gjà aaa	cag Gln	gjà aaa	tgg Trp	cga Arg 215	ctc Leu	ctc Leu	gcg Ala	cct Pro	att Ile 220	acg Thr	gcc Ala	tac Tyr	tcc Ser	2473
caa Gln 225	cag Gln	acg Thr	cga Arg	ggc Gly	cta Leu 230	ctt Leu	ggc Gly	tgc Cys	atc Ile	atc Ile 235	acc Thr	agc Ser	ctc Leu	aca Thr	ggc Gly 240	2521
cgg Arg	gac Asp	agg Arg	aac Asn	cag Gln 245	gtc Val	gag Glu	eja aaa	gag Glu	gtc Val 250	caa Gln	gtg Val	gtc Val	tcc Ser	acc Thr 255	gca Ala	2569
aca Thr	caa Gln	tct Ser	ttc Phe 260	ctg Leu	g <b>cg</b> Ala	acc Thr	tgc Cys	gtc Val 265	aat Asn	ggc	gtg Val	tgt Cys	tgg Trp 270	act Thr	gtc Val	2617
tat Tyr	cat His	ggt Gly 275	gcc Ala	gly ggc	tca Ser	aag Lys	acc Thr 280	ctt Leu	gcc Ala	ggc Gly	cca Pro	aag Lys 285	ggc Gly	cca Pro	atc Ile	2665
acc Thr	caa Gln 290	atg Met	tac Tyr	acc Thr	aat Asn	gtg Val 295	gac Asp	cag Gln	gac Asp	ctc Leu	gtc Val 300	ggc	tgg Trp	caa Gln	gcg Ala	2713
ccc Pro 305	ccc Pro	ejå aaa	gcg Ala	cgt Arg	tcc Ser 310	ttg Leu	aca Thr	cca Pro	tgc Cys	acc Thr 315	tgc Cys	ggc	agc Ser	tcg Ser	gac Asp 320	2761
ctt Leu	tac Tyr	ttg Leu	gtc Val	acg Thr 325	aag Lys	cat His	gcc Ala	gat Asp	gtc Val 330	att Ile	ccg Pro	gtg Val	cgc Arg	cgg Arg 335	cgg Arg	2809
ggc	gac Asp	agc Ser	agg Arg 340	gly aaa	agc Ser	cta Leu	ctc Leu	tcc Ser 345	ccc Pro	cgg Arg	ccc Pro	gtc Val	tcc Ser 350	tac Tyr	ttg Leu	2857
aag Lys	ggc	tct Ser 355	tcg Ser	ggc Gly	ggt Gly	cca Pro	ctg Leu 360	ctc Leu	tgc Cys	ccc Pro	tcg Ser	365 365	cac His	gct Ala	gtg Val	2905
ggc	atc Ile 370	ttt Phe	cgg Arg	gct Ala	gcc Ala	gtg Val 375	tgc Cys	acc Thr	cga Arg	ggg ggg	gtt Val 380	gcg Ala	aag Lys	gcg Ala	gtg Val	2953
gac Asp 385	Phe	gta Val	ccc Pro	gtc Val	gag Glu 390	tct Ser	atg Met	gaa Glu	acc Thr	act Thr 395	atg Met	cgg Arg	tcc Ser	ccg Pro	gtc Val 400	3001
ttc Phe	acg Thr	gac Asp	aac Asn	tcg Ser 405	tcc Ser	cct Pro	ccg Pro	gcc Ala	gta Val 410	ccg Pro	cag Gln	aca Thr	ttc Phe	cag Gln 415	gtg Val	3049
gcc Ala	cat His	cta Leu	cac His	gcc Ala	cct Pro	act Thr	ggt Gly	agc Ser 425	ggc	aag Lys	agc Ser	act Thr	aag Lys 430	gtg Val	ccg Pro	3097

									gtg Val							3145
									tat Tyr							3193
									agg Arg							3241
									ttt Phe 490							3289
					_				tgt Cys	-		_				3337
_	_				_				aca Thr	-	_					3385
_	_			_		_			gcc Ala		_	_		_		3433
tag	atc	acc	ata	cca	cat	cca	aac	atc	αaσ	gag	gtg	gct	ctq	tcc	age	3481
Ser 545									Glu							
545 act	Val gga	Thr gaa	Val	Pro	His 550 ttt	Pro tat	Asn ggc	Ile aaa		Glu 555 atc	Val	Ala	Leu	Ser	Ser 560 atc	3529
545 act Thr	gga gly	Thr gaa Glu ggg	Val atc Ile	Pro ccc Pro 565	His 550 ttt Phe	Pro tat Tyr	Asn ggc Gly	Ile aaa Lys tgc	Glu gcc Ala	Glu 555 atc Ile	Val ccc Pro	Ala atc Ile	Leu gag Glu aaa	acc Thr 575	Ser 560 atc Ile gat	3529 3577
act Thr aag Lys	Val gga Gly ggg Gly	Thr gaa Glu 999 Gly gcc	val atc Ile agg Arg 580	Pro ccc Pro 565 cac His	His 550 ttt Phe ctc Leu	tat Tyr att Ile	ggc Gly ttc Phe	aaa Lys tgc Cys 585	Glu gcc Ala 570 cat	Glu 555 atc Ile tcc Ser	val ccc pro aag Lys	Ala atc Ile aag Lys	gag Glu aaa Lys 590	acc Thr 575 tgc Cys	ser 560 atc Ile gat Asp	
act Thr aag Lys gag Glu	Val gga Gly ggg Gly ctc Leu	Thr gaa Glu 999 Gly gcc Ala 595	val atc Ile agg Arg 580 gcg Ala	Pro ccc Pro 565 cac His aag Lys	His 550 ttt Phe ctc Leu ctg Leu	tat Tyr att Ile tcc ser	ggc Gly ttc Phe ggc Gly 600	aaa Lys tgc Cys 585 ctc Leu	Glu gcc Ala 570 cat His	Glu 555 atc Ile tcc Ser ctc Leu	ccc Pro aag Lys aat Asn	Ala atc Ile aag Lys gct Ala 605	gag Glu aaa Lys 590 gta Val	acc Thr 575 tgc Cys gca Ala	ser 560 atc Ile gat Asp tat Tyr	3577
act Thr aag Lys gag Glu tac Tyr	Val gga Gly ggg Gly ctc Leu cgg Arg 610	Thr gaa Glu ggg Gly gcc Ala 595 ggc Gly	atc Ile agg Arg 580 gcg Ala ctt Leu acg	Pro ccc Pro 565 cac His aag Lys gat Asp	His 550 ttt Phe ctc Leu ctg Leu gta Val	Pro tat Tyr att Ile tcc Ser tcc Ser 615	Asn ggc Gly ttc Phe . ggc Gly 6000 gtc Val	aaa Lys tgc Cys 585 ctc Leu ata Ile	Glu gcc Ala 570 cat His	Glu 555 atc Ile tcc Ser ctc Leu act Thr	Val ccc Pro aag Lys aat Asn agc Ser 620 acc	atc Ile aag Lys gct Ala 605 gga Gly	gag Glu aaaa Lys 590 gta Val gac Asp	acc Thr 575 tgc Cys gca Ala gtc Val	ser 560 atc Ile gat Asp tat Tyr att Ile	3577 3625

				ttc Phe												3817
				cag Gln												3865
				gtg Val												3913
				tgc Cys												3961
	_		_	gag Glu 725			_		-		_				aca Thr	4009
				gtc Val												4057
				acc Thr			_						_		aag Lys :	4105
				aac Asn												4153
				cag Gln												4201
_				cta Leu 805	_		_	_				_				4249
				gcc Ala												4297
				atg Met	_	_					-					4345
				ctg Leu												4393
tgc Cys 865	ctg Leu	aca Thr	aca Thr	ggc Gly	agc Ser 870	gtg Val	gtc Val	att Ile	gtg Val	ggc Gly 875	agg Arg	atc Ile	atc Ile	ttg Leu	tcc Ser 880	4441

#### 64 / 93

.

gga agg cgg gcc atc att ccc gac agg gaa gtc ctt tac cgg gag ttc description of the set o									•								
Asp Glu Met Glu Glu Cys Ala Ser His Leu Pro Tyr Ile Glu Gln Gly 900 905 4585 4585 4585 4585 4585 4585 4585 45			-		Ile			_		Glu	-				Glu		4489
Met Gln Leu Ala Glu Gln Phe Lys Gln Lys Ala Ile Gly Leu Leu Gln 915  aca gcc acc aag caa gcg gag gct gct gct ccc gtg gtg gaa tcc aag 17 Ala Thr Lys Gln Ala Glu Ala Ala Ala Pro Val Val Glu Ser Lys 930  tgg cgg acc ctc gaa gcc ttc tgg gcg aag cat atg tgg aat ttc atc 17 Ala Thr Lys Gln Ala Phe Trp Ala Lys His Met Trp Asn Phe Ile 945  agc ggg ata caa tat tta gca ggc ttg tcc act ctg cct ggc aac ccc 17 Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro 965  gcg ata gca tca ctg atg gca ttc aca gcc tct atc acc agc ccg ctc Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser Pro Leu 980  acc acc caa cat acc ctc ctg ttt acc atc ctg ggg gga tgg gtg gcc thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala 1005  gcc caa ctt gct cct ccc agc gct gct tcc gct ttc gta ggc gcc ggc Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe Val Gly Ala Gly 1010  atc gct gga gcg gct gtt ggc agc ata ggc ctt ggg aag gtg ctt gtg Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly Lys Val Leu Val 1025  atc gct gga gcg gct gtt ggc agc agc ggg gg gg gg gg gg gg gc gg gg at att ttg gca gg				Glu					His					Glu			4537
thr Ala Thr Lys Gln Ala Glu Ala Ala Ala Pro Val Val Glu Ser Lys 930    tgg cgg acc ctc gaa gcc ttc tgg gcg aag cat atg tgg aat ttc atc 194    sys			Leu					Lys					Gly				4585
Trp Arg Thr Leu Glu Ala Phe Trp Ala Lys His Met Trp Asn Phe Ile 945 950 960 975 975 976 976 976 977 975 975 975 975 975 975 975 975 975		Ala					Glu					Val					4633
Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro 975  gcg ata gca tca ctg atg gca ttc aca gcc tct atc acc agc ccg ctc 4777  Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser Pro Leu 980  acc acc caa cat acc ctc ctg ttt aac atc ctg ggg gga tgg gtg gcc 4825  Thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala 995  gcc caa ctt gct cct ccc agc gct gct tcc gct ttc gta ggc gcc ggc 4873  Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe Val Gly Ala Gly 1010  atc gct gga gcg gct gtt ggc agc ata ggc ctt ggg aag gtg ctt gtg Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly Lys Val Leu Val 1025  gat att ttg gca ggt tat gga gca ggg gtg gca ggc gcg ctc gtg gcc Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala 1055  ttt aag gtc atg agc ggc gaf atg ccc tcc acc gag gac ctg gtt aac 5017  Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val Ala 1075  ttt ccc cct gct atc ctc tcc cct gc gcc cta gtc gtc ggg gtc gtg gcc gcg gcg ctc gtg gcc ctc gtg gcc ctc ct	$\operatorname{Trp}$					Ala				_	His	_				Ile	4681
Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser Pro Leu 980 985 985 980 990 4825  acc acc caa cat acc ctc ctg ttt aac atc ctg ggg gga tgg gtg gcc 4825  Thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala 1000 1005  gcc caa ctt gct cct ccc agc gct gct tcc gct ttc gta ggc gcc ggc 4873  Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe Val Gly Ala Gly 1010 1015 1020  atc gct gga gcg gct gtt ggc agc ata ggc ctt ggg aag gtg ctt gtg 11e Ala Gly Ala Gly Ser Ile Gly Leu Gly Lys Val Leu Val 1025 1030 1035 1040  gat att ttg gca ggt tat gga gca ggg gtg gt gta ggc ggc gcg ctc gtg gcc Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala 1045 1050 1055  ttt aag gtc atg agc ggc gag atg ccc tcc acc gag gac ctg gtt aac 5017  Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val Asn 1060 1065  cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc gtg gtg 1065  Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val 1085  tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg gct gtg 5113  Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val	-				Tyr		_		_	ser		_			Asn		4729
Thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala 995  gcc caa ctt gct cct ccc agc gct gct tcc gct ttc gta ggc gcc ggc Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe Val Gly Ala Gly 1010  atc gct gga gcg gct gtt ggc agc ata ggc ctt ggg aag gtg ctt gtg Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly Lys Val Leu Val 1025  gat att ttg gca ggt tat gga gca ggg gtg gca ggc ggg gtg gca ggc gct gtg gcc Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala 1045  ttt aag gtc atg agc ggc gag atg ccc tcc acc gag gac ctg gtt aac Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val Asn 1060  cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc ggg gtc gtg Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val 1085  tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg gct gtg Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val				Ser					Thr					Ser			4777
Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe Val Gly Ala Gly 1010 1015 1020 1020  atc gct gga gcg gct gtt ggc agc ata ggc ctt ggg aag gtg ctt gtg 11e Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly Lys Val Leu Val 1025 1030 1035 1040  gat att ttg gca ggt tat gga gca ggg gtg gca ggc gcg ctc gtg gcc 4969  Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala 1045 1050 1055  ttt aag gtc atg agc ggc gag atg ccc tcc acc gag gac ctg gtt aac 1065 1065  ttt aag gtc atg agc ggc gag atg ccc tcc acc gag gac ctg gtt aac 1060 1065 1070  cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc gtc ggg gtc gtg 1070  cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc gtc ggg gtc gtg 1075  tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg ggg gct gtg 5113  Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val			Gln	His				Phe	Asn				Gly	$\operatorname{Trp}$			4825
Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly Lys Val Leu Val 1025 1030 1035 1040  gat att ttg gca ggt tat gga gca ggg gtg gca ggc gcg ctc gtg gcc 4969  Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala 1045 1050 1055  ttt aag gtc atg agc ggc gag atg ccc tcc acc gag gac ctg gtt aac Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val Asn 1060 1065 1070  cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc gtc ggg gtc gtg Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val 1075 1080 1085  tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg gct gtg S113  Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val	_	Gln	Leu	-			Ser	Ala	_		_	Phe	Val				4873
Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala 1055  ttt aag gtc atg agc ggc gag atg ccc tcc acc gag gac ctg gtt aac Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val Asn 1060  cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc gtc ggg gtc gtg Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val 1075  tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg gct gtg Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val	Ile	Ala				Val	Gly				Leu	Gly				Val	4921
Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val Asn 1060 1065 1070  cta ctc cct gct atc ctc tcc cct ggc gcc cta gtc gtc ggg gtc gtg Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val 1075 1080 1085  tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg gct gtg Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val				Ala	Gly	Tyr				Val	Ala				Val	Ala	4969
Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val 1075  tgc gca gcg ata ctg cgt cgg cac gtg ggc cca ggg gag ggg gct gtg Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val				Met	Ser				Pro	Ser				Leu	Val		5017
Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val			Pro	Ala				Pro	Gly				Val	Gly			5065
		Ala	Ala				Arg	His				Gly	Glu				5113

cag tgg atg aac c Gln Trp Met Asn A 1105	gg ctg ata gcg rg Leu Ile Ala 1110	ttc gct tcg cgg Phe Ala Ser Arg 1115	ggt aac cac Gly Asn His	gtc 5161 Val 1120
tcc ccc acg cac t Ser Pro Thr His T				Thr
cag atc ctc tct a Gln Ile Leu Ser S 1140				
cag tgg atc aac g Gln Trp Ile Asn G 1155	ag gac tgc tcc lu Asp Cys Ser 1160	Thr Pro Cys Ser	ggc tcg tgg Gly Ser Trp 1165	cta 5305 Leu
aga gat gtt tgg g Arg Asp Val Trp A 1170	at tgg ata tgc sp Trp Ile Cys 1175	acg gtg ttg act Thr Val Leu Thr 118	Asp Phe Lys	gcc 5353 Ala
tgg ctc cag tcc a Trp Leu Gln Ser L 1185				
tca tgt caa cgt g Ser Cys Gln Arg G 1	gg tac aag gga ly Tyr Lys Gly .205	gtc tgg cgg ggc Val Trp Arg Gly 1210	gac ggc atc Asp Gly Ile 121	Met
caa acc acc tgc c Gln Thr Thr Cys P 1220		_		
tgt tcc atg agg a Cys Ser Met Arg I 1235	tc gtg ggg cct le Val Gly Pro 1240	Arg Thr Cys Ser	aac acg tgg Asn Thr Trp 1245	cat 5545 His
gga aca ttc ccc a Gly Thr Phe Pro I 1250			Cys Thr Pro	
ccg gcg cca aat t Pro Ala Pro Asn T 1265	at tot agg gcg Yr Ser Arg Ala 1270	ctg tgg cgg gtg Leu Trp Arg Val 1275	g get get gag . Ala Ala Glu	gag 5641 Glu 1280
tac gtg gag gtt a Tyr Val Glu Val T				Met
acc act gac aac g Thr Thr Asp Asn V ; 1300	rta aag tgc ccg /al Lys Cys Pro	tgt cag gtt ccg Cys Gln Val Pro 1305	gcc ccc gaa Ala Pro Glu 1310	ttc 5737 Phe .
ttc aca gaa gtg g Phe Thr Glu Val A 1315		Leu His Arg Tyr		

1330				c ctg gtc e Leu Val 1340	Gly Leu		5833
tac ccg gtt o Tyr Pro Val ( 1345	Gly Ser G						5881
gtg ctc act t Val Leu Thr s	_	_		r His Ile			5929
gct aag cgt a Ala Lys Arg A						Ser Ser	5977
tca gct agc ( Ser Ala Ser ( 1395		Ser Ala Pr	-		-		6025
cgt cat gac t Arg His Asp 8 1410					Asn Leu		6073
cgg cag gag a Arg Gln Glu I 1425	Met Gly G						6121
gta gta att i Val Val Ile I							6169
	1445		149	50		1455	
agg gaa gta f Arg Glu Val S	1445 too gtt o	ccg gcg ga	ag atc cto	g cgg agg	tcc agg	aaa ttc Lys Phe	6217
agg gaa gta f Arg Glu Val S	1445 tcc gtt c Ser Val F 1460 atg ccc a	ccg gcg ga Pro Ala Gl ata tgg gc Ile Trp Al	ag atc cto lu Ile Lev 1465 ca cgc cco	g cgg agg u Arg Arg	tcc agg Ser Arg 1470	aaa ttc Lys Phe ) cca ctg	6217 6265
agg gaa gta f Arg Glu Val S cct cga gcg a Pro Arg Ala N	1445 tcc gtt c Ser Val F 1460 atg ccc a Met Pro I	ccg gcg ga Pro Ala Gl Ata tgg gc Ile Trp Al 14 gac ccg ga	ag atc ctg lu Ile Len 1465 ca cgc ccg la Arg Pro 180	g cgg agg u Arg Arg g gat tac o Asp Tyr	tcc agg Ser Arg 1470 aac cct Asn Pro 1485 gtg gta Val Val	aaa ttc Lys Phe cca ctg Pro Leu	
agg gaa gta fa Arg Glu Val S cct cga gcg a Pro Arg Ala M 1475 tta gag tcc fa	tcc gtt c Ser Val F 1460 atg ccc a Met Pro I tgg aag g Trp Lys A ccg cct g	ccg gcg ga Pro Ala Gl  Ata tgg gc Ile Trp Al  14 gac ccg ga Asp Pro As 1495 gcc aag gc	ag atc ctg lu Ile Let 1465 ca cgc ccg la Arg Pro 180 ac tac gtc sp Tyr Val	g cgg agg u Arg Arg g gat tac o Asp Tyr c cct cca l Pro Pro 1500	tcc agg Ser Arg 1470 aac cct Asn Pro 1485 gtg gta Val Val cct cca	aaa ttc Lys Phe  cca ctg Pro Leu  cac ggg His Gly	6265
agg gaa gta ta Arg Glu Val S cct cga gcg a Pro Arg Ala M 1475 tta gag tcc t Leu Glu Ser 1 1490 tgt cca ttg c Cys Pro Leu I	tcc gtt c Ser Val F 1460  atg ccc a Met Pro I  tgg aag g Trp Lys A  ccg cct g Pro Pro A	ccg gcg ga Pro Ala Gl  ata tgg gc Ile Trp Al	ag atc ctg lu Ile Let 1465  ca cgc ccg la Arg Pro 180  ac tac gtc sp Tyr Val cc cct ccg la Pro Pro ac tct ac	g cgg agg g Arg Arg g gat tac c Asp Tyr c cct cca l Pro Pro 1500 g ata cca o Ile Pro 1515 c gtg tct r Val Ser	tcc agg Ser Arg 1470 aac cct Asn Pro 1485 gtg gta Val Val cct cca Pro Pro	aaa ttc Lys Phe  cca ctg Pro Leu  cac ggg His Gly  cgg agg Arg Arg 1520  ttg gcg	6265 6313

Ser Gly Thr Ala Th		ac cag ccc tcc gac sp Gln Pro Ser Asp 156	Asp Gly Asp
gcg gga tcc gac gt Ala Gly Ser Asp Va 1570			
gag ccg ggg gat cc Glu Pro Gly Asp Pr 1585			
gag gag gct agt ga Glu Glu Ala Ser Gl 16			
aca ggc gcc ctg at Thr Gly Ala Leu Il 1620	e Thr Pro Cys Al		
atc aat gca ctg ag Ile Asn Ala Leu Se 1635			Leu Val Tyr
gct aca aca tct cg Ala Thr Thr Ser Ar 1650			
	c cta asc asc ca		
Asp Arg Leu Gln Va		ac tac cgg gac gtg is Tyr Arg Asp Val 1675	
Asp Arg Leu Gln Va 1665 atg aag gcg aag gc Met Lys Ala Lys Al	l Leu Asp Asp Hi 1670 g tcc aca gtt aa	is Tyr Arg Asp Val 1675 ag get aaa ett eta	Leu Lys Glu 1680 tcc gtg gag 6889
Asp Arg Leu Gln Va 1665 atg aag gcg aag gc Met Lys Ala Lys Al	1 Leu Asp Asp Hi 1670 g tcc aca gtt as a Ser Thr Val Ly 85 g acg ccc cca cs u Thr Pro Pro Hi	is Tyr Arg Asp Val 1675 ag gct aaa ctt cta 78 Ala Lys Leu Leu 1690 at tcg gcc aga tct	Leu Lys Glu 1680  tcc gtg gag 6889 Ser Val Glu 1695  aaa ttt ggc 6937
Asp Arg Leu Gln Va 1665  atg aag gcg aag gc Met Lys Ala Lys Al  16  gaa gcc tgt aag ct Glu Ala Cys Lys Le	1 Leu Asp Asp Hi 1670  g tcc aca gtt aa a Ser Thr Val Ly 85  g acg ccc cca ca u Thr Pro Pro Hi c gtc cgg aac ct	ag gct aaa ctt cta 78 Ala Lys Leu Leu 1690 At tcg gcc aga tct 1s Ser Ala Arg Ser 705	Leu Lys Glu 1680  tcc gtg gag Ser Val Glu 1695  aaa ttt ggc 6937 Lys Phe Gly 1710  gtt aac cac 6985 Val Asn His
Asp Arg Leu Gln Vale 1665  atg aag gcg aag gc Met Lys Ala Lys Al 16  gaa gcc tgt aag ct Glu Ala Cys Lys Le 1700  tat ggg gca aag ga Tyr Gly Ala Lys As	1 Leu Asp Asp Hi 1670  g tcc aca gtt aa a Ser Thr Val Ly 85  g acg ccc cca ca u Thr Pro Pro Hi c gtc cgg aac ct p Val Arg Asn Le 1720  g aag gac ttg ct	ag gct aaa ctt cta 78 Ala Lys Leu Leu 1690  at tcg gcc aga tct 1s Ser Ala Arg Ser 705  a tcc agc aag gcc at tcc agc aag gcc au Ser Ser Lys Ala 172	Leu Lys Glu 1680  tcc gtg gag 6889 Ser Val Glu 1695  aaa ttt ggc 6937 Lys Phe Gly 1710  gtt aac cac 6985 Val Asn His 5
Asp Arg Leu Gln Vale 1665  atg aag gcg aag gc Met Lys Ala Lys Ala Lys Ala Glu Ala Cys Lys Leg 1700  tat ggg gca aag ga Tyr Gly Ala Lys As 1715  atc cgc tcc gtg tg Ile Arg Ser Val Tr	1 Leu Asp Asp Hi 1670  g tcc aca gtt aa a Ser Thr Val Ly 85  g acg ccc cca ca u Thr Pro Pro Hi  c gtc cgg aac ct p Val Arg Asn Le 1720  g aag gac ttg ct p Lys Asp Leu Le 1735  g gca aaa aat ga	ag gct aaa ctt cta 78 Ala Lys Leu Leu 1690  at tcg gcc aga tct 1s Ser Ala Arg Ser 705  at tcc agc aag gcc at tcc agc aag gcc at Ser Ser Lys Ala 172  ag gaa gac act gag au Glu Asp Thr Glu 1740  ag gtt ttc tgc gtc	Leu Lys Glu 1680  tcc gtg gag Ser Val Glu 1695  aaa ttt ggc 6937 Lys Phe Gly 1710  gtt aac cac 6985 Val Asn His 5  aca cca att 7033 Thr Pro Ile  caa cca gag 7081

# 68 / 93

gtt cgt gtg tgc Val Arg Val Cys 1780	Glu Lys Met .				7177
cct cag gcc gtg Pro Gln Ala Val 1795	Met Gly Ser			Ser Pro Gly	7225
cag cgg gtc gag Gln Arg Val Glu 1810		Asn Ala Trp			7273
atg ggc ttc gca Met Gly Phe Ala 1825					7321
aat gac atc cgt Asn Asp Ile Arg			Gln Cys Cys		7369
ccc gaa gcc aga Pro Glu Ala Arg 1860	Gln Ala Ile				7417
ggg ggc ccc ctg Gly Gly Pro Leu 1875	Thr Asn Ser			Tyr Arg Arg	7465
tgc cgc gcg agc Cys Arg Ala Ser 1890		Thr Thr Ser			7513
tgt tac ttg aag Cys Tyr Leu Lys 1905					7561
tgc acg atg ctc Cys Thr Met Leu			Val Val Ile		7609
gcg ggg acc caa Ala Gly Thr Gln 1940	Glu Asp Glu				7657
atg act aga tac Met Thr Arg Tyr 1955	Ser Ala Pro			Pro Glu Tyr	7705
gac ttg gag ttg Asp Leu Glu Leu 1970		Cys Ser Ser			7753
gat gca tct ggc Asp Ala Ser Gly 1985					7801

;

03 / 33		
ccc ctt gcg cgg gct gcg tgg gag aca gct aga cac act cca gtcPro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val20052010	Asn	<b>)</b>
tcc tgg cta ggc aac atc atc atg tat gcg ccc acc ttg tgg gca Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro Thr Leu Trp Ala 2020 2025 2030		7
atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu 2035 2040 2045		5
ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser 2050 2055 2060		3
gag cca ctt gac cta cct cag atc att caa cga ctc cac ggc ctt Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu 2065 2070 2075		L
gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val 2085 2090 2099	Ala	Э
tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg 2100 2105 2110		7
cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg 2115 2120 2125		5
gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys 2130 2135 2140		3
aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser 2145		L
ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser 2165 2170 2175	Arg.	Э
gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Leu Ser Val 2180 2185 2190		7
gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactco Val Gly Ile Tyr Leu Leu Pro Asn Arg * 2195. 2200	ca 8427	7
ggccaatagg ccatcetgtt tttttccctt ttttttttc tttttttt ttttttt	tggct 8547	7 7

<211> 6 <212> DNA <213> HCV						
<400> 8 accagc					•	6
<210> 9 <211> 63 <212> DNA <213> HCV						
<400> 9 gaattccaga gac	tggcgcgccc	agatgttaac	cagatccatg	gcacactcta	gagtactgtc	60 63
<210> 10 <211> 33 <212> DNA <213> HCV						
<400> 10 cggaatcgtt	aacagaccac	aacggtttcc	ctc			33
<210> 11 <211> 30 <212> DNA <213> HCV						
<400> 11 ggcgtaccca	tggtattatc	gtgtttttca				30
<210> 12 <211> 45 <212> DNA <213> HCV						
<400> 12 gcatatgaat	tctaatacga	ctcactatag	gccagccccc	gattg		45
<210> 13 <211> 45 <212> DNA <213> HCV						
<400> 13 ggcgcgccct	ttggtttttc	tttgaggttt	aggattcgtg	ctcat		45
<210> 14 <211> 36 <212> DNA <213> HCV						
<400> 14 aaagggcgca	tgattgaaca	agatggattg	cacgca			36
<210> 15 <211> 39						

<212> DNA <213> HCV	
<400> 15 gcatatgtta actcagaaga actcgtcaag aaggcgata 3	9
<210> 16 <211> 45 <212> DNA <213> HCV	
<400> 16 gcatatgaat totaatacga otcactatag gccagcocco gattg 4	١5
<210> 17 <211> 30 <212> DNA <213> HCV	
<400> 17 acgcagaaag cgtctagcca tggcgttagt .	0
<210> 18 <211> 30 <212> DNA <213> HCV	
<400> 18 tcccggggca ctcgcaagca ccctatcagg 3	0
<210> 19 <211> 26 <212> DNA <213> HCV	
<220> <223> Label with FAM: fluorescence reporter dye	
<223> Label with TAMRA: Quencher dye	
<400> 19 tggtctgcgg aacgggtgag tacacc 2	6
<210> 20 <211> 45 <212> DNA <213> HCV	
<400> 20 gtggacgaat totaatacga otcactataa ccagooccog attgg 4	5
<210> 21 <211> 27 <212> DNA <213> HCV	
<400> 21 ggaacgcccg tcgtggccag ccacgat 2	7

```
<210> 22
<211> 23
<212> DNA
<213> HCV
<400> 22
gtcgtcttct ctgacatgga gac
                                                                   23
<210> 23
<211> 27
<212> DNA
<213> HCV
<400> 23
gagttgctca gtggattgat gggcagc
                                                                   27
<210> 24
<211> 8638
<212> DNA
<213> HCV
<220>
<221> CDS
<222> (1802)...(8407)
<400> 24
accagecece gattggggge gacactecae catagateae teceetqtga qqaactaetq 60
tetteaegea gaaagegtet agecatggeg ttagtatgag tgtegtgeag eetceaggae 120
cccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaccggg teetttettg gatcaacccg etcaatgeet ggagatttgg gegtgeecce 240
gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
eggeegettg ggtggagagg etattegget atgactggge acaacagaca ateggetget 480
ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgacgggcgt tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtateeat catggetgat geaatgegge ggetgeatae gettgateeg getaeetgee 780
cattegacea ceaagegaaa categeateg agegageaeg taeteggatg gaageeggte 840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900
ccaggctcaa ggcgcgcatg cccgacggcg aggatctcgt cgtgacccat ggcgatgcct 960
gettgeegaa tateatggtg gaaaatggee gettttetgg atteategae tgtggeegge 1020
tgggtgtggg ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggcggcga atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc 1140
agegeatege ettetatege ettettgaeg agttettetg agttegegee cagatgttaa 1200
cagaccacaa eggttteeet etagegggat caatteegee eeececeta aegttactgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
gccgtctttt ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc 1380
taggggtett teceeteteg ecaaaggaat geaaggtetg ttgaatgteg tgaaggaage 1440
agtteetetg gaagettett gaagacaaac aacgtetgta gegaceettt geaggeageg 1500
gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560
tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggctctcc tcaagcgtat tcaacaaggg gctgaaggat gcccagaagg taccccattg 1680
tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaaa 1740
aaacgtetag geeceegaa eeaeggggae gtggttttee tttgaaaaac acgataatac 1800
c atg gac egg gag atg gca gca teg tge gga gge geg gtt tte gta ggt 1849
```

Met Asp Arg Glu Met A		ly Gly Ala Val Ph LO	e Val Gly 15
ctg ata ctc ttg acc ttg Leu Ile Leu Leu Thr Leu 20	tca ccg cac tat Ser Pro His Tyr 25	aag ctg ttc ctc Lys Leu Phe Leu 30	gct agg 1897 Ala Arg
ctc ata tgg tgg tta cas Leu Ile Trp Trp Leu Glr 35	tat ttt atc acc Tyr Phe Ile Thr 40	agg gcc gag gca Arg Ala Glu Ala : 45	cac ttg 1945 His Leu
caa gtg tgg atc ccc ccc Gln Val Trp Ile Pro Pro 50	ctc aac gtt cgg Leu Asn Val Arg 55	ggg ggc cgc gat Gly Gly Arg Asp 60	gcc gtc 1993 Ala Val
atc ctc ctc acg tgc gcc Ile Leu Leu Thr Cys Ala 65	l Ile His Pro Glu	cta atc ttt acc Leu Ile Phe Thr 75	atc acc 2041 Ile Thr 80
aaa atc ttg ctc gcc ata Lys Ile Leu Leu Ala Ile 85			
ata acc aaa gtg ccg tad Ile Thr Lys Val Pro Tyn 100			
tgc atg ctg gtg cgg aag Cys Met Leu Val Arg Lys 115	g gtt got ggg ggt : Val Ala Gly Gly 120	cat tat gtc caa His Tyr Val Gln 125	atg gct 2185 Met Ala
ctc atg aag ttg gcc gca Leu Met Lys Leu Ala Ala 130	ctg aca ggt acg Leu Thr Gly Thr 135	tac gtt tat gac Tyr Val Tyr Asp 140	cat ctc 2233 His Leu
acc cca ctg cgg gac tgg Thr Pro Leu Arg Asp Trp 145	Ala His Ala Gly	cta cga gac ctt Leu Arg Asp Leu 155	gcg gtg 2281 Ala Val 160
gca gtt gag ccc gtc gtc Ala Val Glu Pro Val Val 165		Glu Thr Lys Val	
tgg ggg gca gac acc gcg Trp Gly Ala Asp Thr Ala 180	g gcg tgt ggg gac Ala Cys Gly Asp 185	atc atc ttg ggc Ile Ile Leu Gly 190	ctg ccc 2377 Leu Pro
gte tee gee ege agg ggg Val Ser Ala Arg Arg Glg 195	g agg gag ata cat Arg Glu Ile His 200	ctg gga ccg gca Leu Gly Pro Ala 205	gac agc 2425 Asp Ser
ctt gaa ggg cag ggg tgg Leu Glu Gly Gln Gly Trp 210	g cga ctc ctc gcg Arg Leu Leu Ala 215	cct att acg gcc Pro Ile Thr Ala 220	tac tcc 2473 Tyr Ser
caa cag acg cga ggc cta Gln Gln Thr Arg Gly Len 225	Leu Gly Cys Ile	atc act agc ctc Ile Thr Ser Leu 235	aca ggc 2521 Thr Gly 240

cgg Arg	gac Asp	agg Arg	aac Asn	cag Gln 245	gtc Val	gag Glu	gjà aaa	gag Glu	gtc Val 250	caa Gln	gtg Val	gtc Val	tcc Ser	acc Thr 255	gca Ala	2569
											gtg Val					2617
											cca Pro					2665
											gtc Val 300					2713
											tgc Cys					2761
											ccg Pro					2809
ggc ggc	gac Asp	agc Ser	agg Arg 340	eja aaa	agc Ser	cta Leu	ctc Leu	tcc Ser 345	ccc Pro	agg Arg	ccc Pro	gtc Val	tcc Ser 350	tac Tyr	ttg Leu	2857
aag Lys	gly ggc	tct Ser 355	tcg Ser	ggc	ggt Gly	cca Pro	ctg Leu 360	ctc Leu	tgc Cya	ccc Pro	tcg Ser	365 Gly 999	cac His	gct Ala	gtg Val	2905
											gtt Val 380					2953
											atg Met					3001
											cag Gln					3049
											agc Ser				_	3097
											gtc Val					3145
_	_	_								_	tct Ser 460	_				3193

	gac Asp															3241
	atc Ile															3289
	elà aaa															3337
	tcg Ser															3385
_	gct Ala 530			_		_	-		_					-		3433
-	gtc Val											-	_		-	3481
	gga Gly	_							_							3529
-	<b>el</b> aaa	~~~									_	-		_	_	3577
	ctc Leu	-		_	_							_				3625
	cgg Arg 610															3673
-	gta Val	-	-	-	-		-	-						_		3721
	gtg Val		_	_						_						3769
	gac Asp															3817
	tca Ser															3865

								gaa Glu								3913
								gac Asp								3961
								agg Arg								4009
cca Pro	gly aaa	ttg Leu	ccc Pro 740	gtc Val	tgc Cys	cag Gln	gac Asp	cat His 745	ctg Leu	gag Glu	ttc Phe	tgg Trp	gag Glu 750	agc Ser	gtc Val	4057
								gcc Ala								4105
								ctg Leu								4153
								cca Pro								4201
								ctg Leu								4249
								gag Glu 825								4297
acc Thr	aaa Lys	tac Tyr 835	atc Ile	atg Met	gca Ala	tgc Cys	atg Met 840	tcg Ser	gct Ala	gac Asp	ctg Leu	gag Glu 845	gtc Val	gtc Val	acg Thr	4345
								gtc Val								4393
tgc Cys 865	ctg Leu	aca Thr	aca Thr	ggc	agc Ser 870	gtg Val	gtc Val	att Ile	gtg Val	ggc Gly 875	agg Arg	atc Ile	atc Ile	ttg Leu	tcc Ser 880	4441
								agg Arg								4489
gat Asp	gag Glu	atg Met	gaa Glu 900	gag Glu	Cya tgc	gcc Ala	tca Ser	cac His 905	ctc Leu	cct Pro	tac Tyr	atc Ile	gaa Glu 910	cag Gln	gga Gly	4537

Met Gln Leu A 915	-	-	aag gca atc Lys Ala Ile		
aca gcc acc a Thr Ala Thr L 930					
tgg cgg acc c Trp Arg Thr L 945					
agc gġg ata c Ser Gly Ile G					
gcg ata gca to Ala Ile Ala So 99					
acc acc caa co Thr Thr Gln H					
gcc caa ctt g Ala Gln Leu A 1010				Val Gly	
ata aat aan a	ra act att	ggg agg ata	aga att aga	aan ata	ctt ata 4921
atc gct gga go Ile Ala Gly A 1025	la Ala Val 1030	Gly Ser Ile	Gly Leu Gly 1035	Lys Val	Leu Val 1040
Ile Ala Gly A	la Ala Val 1030 ca ggt tat	Gly Ser Ile gga gca ggg	Gly Leu Gly 1035 gtg gca ggc	Lys Val gcg ctc Ala Leu	Leu Val 1040 gtg gcc 4969
Ile Ala Gly Al 1025 gat att ttg ga Asp Ile Leu Al ttt aag gtc at Phe Lys Val Me	la Ala Val 1030 ca ggt tat la Gly Tyr 1045 cg agc ggc	Gly Ser Ile  gga gca ggg Gly Ala Gly  gag atg ccc	Gly Leu Gly 1035 gtg gca ggc Val Ala Gly 1050 tcc acc gag Ser Thr Glu	Lys Val  gcg ctc Ala Leu  gac ctg	Leu Val 1040  gtg gcc 4969 Val Ala 1055  gtt aac 5017 Val Asn
Ile Ala Gly Al 1025 gat att ttg ga Asp Ile Leu Al ttt aag gtc at Phe Lys Val Me	La Ala Val 1030  Ca ggt tat La Gly Tyr 1045  Cg agc ggc et Ser Gly 060  Ct atc ctc	gga gca ggg Gly Ala Gly gag atg ccc Glu Met Pro 1069	Gly Leu Gly 1035  gtg gca ggc Val Ala Gly 1050  tcc acc gag Ser Thr Glu 5	gcg ctc Ala Leu gac ctg Asp Leu 1070 gtc ggg	Leu Val 1040 gtg gcc 4969 Val Ala 1055 gtt aac 5017 Val Asn
Ile Ala Gly Al 1025  gat att ttg gat att ttt aag gtc at Phe Lys Val Market val Market gat att att att att att att att att at	La Ala Val 1030  Ca ggt tat La Gly Tyr 1045  Cg agc ggc et Ser Gly 060  Ct atc ctc La Ile Leu  La ctg cgt Le Leu Arg	gga gca ggg Gly Ala Gly gag atg ccc Glu Met Pro 1069 tcc cct ggc Ser Pro Gly 1080 cgg cac gtg	Gly Leu Gly 1035  gtg gca ggc Val Ala Gly 1050  tcc acc gag Ser Thr Glu 5  gcc cta gtc Ala Leu Val	gcg ctc Ala Leu  gac ctg Asp Leu 1070 gtc ggg Val Gly 1085 gag ggg Glu Gly	Leu Val 1040  gtg gcc 4969 Val Ala 1055  gtt aac 5017 Val Asn  gtc gtg 5065 Val Val  gct gtg 5113
Ile Ala Gly Al 1025  gat att ttg gat att ttt aag gtc at Phe Lys Val Market att att att att att att att att att a	ta Ala Val 1030 ca ggt tat la Gly Tyr 1045 cg agc ggc et Ser Gly 060 ct atc ctc la Ile Leu la ctg cgt le Leu Arg	gga gca ggg Gly Ala Gly  gag atg ccc Glu Met Pro 1069  tcc cct ggc Ser Pro Gly 1080  cgg cac gtg Arg His Val 1095  ata gcg ttc Ile Ala Phe	gtg gca ggc Val Ala Gly 1050  tcc acc gag Ser Thr Glu  gcc cta gtc Ala Leu Val  ggc cca ggg Gly Pro Gly 1100 gct tcg cgg	gcg ctc Ala Leu  gac ctg Asp Leu 1070  gtc ggg Val Gly 1085  gag ggg Glu Gly	Leu Val

cag atc ctc tct agt ctt acc atc act cag ctg ctg aag agg ctt cac Gln Ile Leu Ser Ser Leu Thr Ile Thr Gln Leu Leu Lys Arg Leu His 1140 1145 1150	5257
cag tgg atc aac gag gac tgc tcc acg cca tgc tcc ggc tcg tgg cta Gln Trp Ile Asn Glu Asp Cys Ser Thr Pro Cys Ser Gly Ser Trp Leu 1155 1160 1165	5305
aga gat gtt tgg gat tgg ata tgc acg gtg ttg act gat ttc aag acc Arg Asp Val Trp Asp Trp Ile Cys Thr Val Leu Thr Asp Phe Lys Thr 1170 1175 1180	5353
tgg ctc cag tcc aag ctc ctg ccg cga ttg ccg gga gtc ccc ttc ttc Trp Leu Gln Ser Lys Leu Leu Pro Arg Leu Pro Gly Val Pro Phe Phe 1185 1190 1195 1200	5401
tea tgt caa egt ggg tac aag gga gte tgg egg gge gae gge ate atg Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile Met 1205 1210 1215	5449
caa acc acc tgc cca tgt gga gca cag atc acc gga cat gtg aaa aac Gln Thr Thr Cys Pro Cys Gly Ala Gln Ile Thr Gly His Val Lys Asn 1220 1225 1230	5497
ggt tcc atg agg atc gtg ggg cct agg acc tgt agt aac acg tgg cat Gly Ser Met Arg Ile Val Gly Pro Arg Thr Cys Ser Asn Thr Trp His 1235 1240 1245	5545
gga aca ttc ccc att aac gcg tac acc acg ggc ccc tgc acg ccc tcc Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro Ser 1250 1255 1260	5593
ccg gcg cca aat tat tct agg gcg ctg tgg cgg gtg gct gct gag gagPro Ala Pro Asn Tyr Ser Arg Ala Leu Trp Arg Val Ala Ala Glu Glu126512701280	5641
tac gtg gag gtt acg cgg gtg ggg gat ttc cac tac gtg acg ggc atg Tyr Val Glu Val Thr Arg Val Gly Asp Phe His Tyr Val Thr Gly Met 1285 1290 1295	5689
acc act gac aac gta aag tgc ccg tgt cag gtt ccg gcc ccc gaa ttc Thr Thr Asp Asn Val Lys Cys Pro Cys Gln Val Pro Ala Pro Glu Phe 1300 1305 1310	5737
ttc aca gaa gtg gat ggg gtg cgg ttg cac agg tac gct cca gcg tgc Phe Thr Glu Val Asp Gly Val Arg Leu His Arg Tyr Ala Pro Ala Cys 1315 1320 1325	5785
aaa ccc ctc cta cgg gag gag gtc aca ttc ctg gtc ggg ctc aat caa Lys Pro Leu Leu Arg Glu Glu Val Thr Phe Leu Val Gly Leu Asn Gln 1330 1335 1340	5833
tac ctg gtt ggg tca cag ctc cca tgc gag ccc gaa ccg gac gta gca Tyr Leu Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val Ala 1345 1350 1355 1360	5881

gtg ctc act tcc atg Val Leu Thr Ser Met 136	Leu Thr Asp Pro			29
gct aag cgt agg ctg Ala Lys Arg Arg Leu 1380	gcc agg gga tct Ala Arg Gly Ser 138	Pro Pro Ser L	tg gcc agc tca 597 eu Ala Ser Ser 1390	17
tca gct agc cag ctg Ser Ala Ser Gln Leu 1395		Leu Lys Ala T		35
cgt cat gac tcc ccg Arg His Asp Ser Pro 1410				73
cgg cag gag atg ggc Arg Gln Glu Met Gly 1425	ggg aac atc acc Gly Asn Ile Thr 1430	cgc gtg gag to Arg Val Glu So 1435	ca gaa aat aag 612 er Glu Asn Lys 1440	31
gta gta att ttg gad Val Val Ile Leu Asp .144	Ser Phe Glu Pro			59
agg gaa gta too gtt Arg Glu Val Ser Val 1460		Leu Arg Arg S		
cct cga gcg atg ccc Pro Arg Ala Met Pro 1475	ata tgg gca cgc Ile Trp Ala Arg 1480	Pro Asp Tyr A	ac cct cca ctg 626 sn Pro Pro Leu 485	55
tta gag tcc tgg aag Leu Glu Ser Trp Lys 1490				L3
tgt cca ttg ccg cct Cys Pro Leu Pro Pro 1505				51
aag agg acg gtt gtc Lys Arg Thr Val Val 152	Leu Ser Glu Ser		-	)9
gag ctc gcc aca aag Glu Leu Ala Thr Lys 1540		Ser Glu Ser S		57
agc ggc acg gca acg Ser Gly Thr Ala Thr 1555	- ·	Gln Pro Ser A	<del>-</del>	)5
gcg gga tcc gac gtt Ala Gly Ser Asp Val				53

		ctc agc gac Leu Ser Asp 0			
gag gag gct Glu Glu Ala	agt gag gad Ser Glu Asp 1605	gtc gtc tgc Val Val Cys	tgc tcg atg Cys Ser Met 1610	tcc tac aca Ser Tyr Thr 161	Trp
aca ggc gcc Thr Gly Ala	ctg atc acc Leu Ile Thr 1620	cca tgc gct Pro Cys Ala 162	Ala Glu Glu	acc aag ctg Thr Lys Leu 1630	ccc 6697 Pro
	Leu Ser Asr	tct ttg ctc Ser Leu Leu 1640			
		gca agc ctg Ala Ser Leu 1655		Lys Val Thr	
		gac gac cac Asp Asp His O			
		aca gtt aag Thr Val Lys			Glu
gaa gcc tgt Glu Ala Cys	aag ctg acg Lys Leu Thi 1700	ccc cca cat Pro Pro His 170	Ser Ala Arg	tct aaa ttt Ser Lys Phe 1710	ggc 6937 Gly
tat ggg gca Tyr Gly Ala 171!	Lys Asp Val	cgg aac cta Arg Asn Leu 1720	tcc agc aag Ser Ser Lys	gcc gtt aac Ala Val Asn 1725	cac 6985 His
atc cgc tcc Ile Arg Ser 1730	gtg tgg aag Val Trp Lys	gac ttg ctg Asp Leu Leu 1735	gaa gac act Glu Asp Thr 174	Glu Thr Pro	att 7033 Ile
		aaa aat gag Lys Asn Glu O			
		gct cgc ctt Ala Arg Leu			Gly
		atg gcc ctt Met Ala Leu 178	Tyr Asp Val		
cct cag gcc				+-a +a+ aa+	gga 7225

	gag ttc ctg Glu Phe Leu			Lys Lys Cys	
	gca tat gac Ala Tyr Asp 1830	Thr Arg Cys			
_	cgt gtt gag Arg Val Glu 1845		-		Ala
	aga cag gcc Arg Gln Ala 1860		Leu Thr Glu		
	ctg act aat Leu Thr Asn 5				
	agc ggt gta Ser Gly Val			Asn Thr Leu	
	aag gcc gct Lys Ala Ala 1910	Ala Ala Cys			
	ctc gta tgc Leu Val Cys 1925		~ ~		Ser
	caa gag gac Gln Glu Asp 1940		Leu Arg Ala		
	tac tct gcc Tyr Ser Ala 5				
	ttg ata aca Leu Ile Thr	_		Ser Val Ala	
I I	ggc aaa agg Gly Lys Arg 1990	Val Tyr Tyr	, _		
	cgg gct gcg Arg Ala Ala 2005				Asn
	ggc aac atc Gly Asn Ile 2020	_	Ala Pro Thr		

atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa caa  Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln  2035  2040  2045	45
ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc att 79. Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile 2050 2055 2060	93
gag cca ctt gac cta cct cag atc att caa cga ctc cac ggc ctt agc 80. Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser 2065 2070 2075 2080	41
gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg gct 80 Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala 2085 2090 2095	89
tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat 81 Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His 2100 2105 2110	.37
cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct 81 Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala 2115 2120 2125	.85
gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc 82 Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu 2130 2135 2140	33
aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg 82 Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp 2145 2150 2155 2160	281
ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt 83 Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg 2165 2170 2175	329
gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta ggg 83 Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly 2180 2185 2190	377
gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca 84 Val Gly Ile Tyr Leu Leu Pro Asn Arg * 2195 2200	127
ggccaatagg ccatcctgtt tttttccctt ttttttttt ttttttttt tttttt	547
<210> 25 <211> 8638 <212> DNA <213> HCV	
<220> <221> CDS <222> (1802)(8407)	

```
<400> 25
accagococo gattgggggo gacactocac catagatcac toccotgtga ggaactactg 60
tetteaegea gaaagegtet ageeatggeg ttagtatgag tgtegtgeag cetecaqgae 120
cocccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gaegaeeggg teetttettg gateaaeeeg eteaatgeet ggagatttgg gegtgeeeee 240
gegagaetge tageegagta gtgttgggte gegaaaggee ttgtggtaet geetgatagg 300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
cggccgcttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct 480
ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaaqaccg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgacgggcgt tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccqaga 720
aagtateeat eatggetgat geaatgegge ggetgeatae gettgateeg getaeetgee 780
cattegacca ccaagegaaa categeateg agegageaeg tacteggatg gaageeggte 840
tigicgatca ggatgatcig gacgaagagc atcaggggct cgcgccagcc gaactgitcg 900
ccaggeteaa ggegegeatg ccegaeggeg aggatetegt egtgaeceat ggegatgeet 960
gettgeegaa tateatggtg gaaaatggee gettttetgg atteategae tgtggeegge 1020
tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggeggega atgggetgae egetteeteg tgetttaegg tategeeget eeegattege 1140
agegeatege ettetatege ettettgaeg agttettetg agttegegee eagatgttaa 1200
cagaccacaa cggtttccct ctagcgggat caattccgcc cccccccta acgttactgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
geogtetttt ggeaatgtga gggeeeggaa acetggeeet gtettettga egageattee 1380
taggggtett teceeteteg ecaaaggaat geaaggtetg ttgaatgteg tgaaggaage 1440
agtteetetg gaagettett gaagacaaac aacgtetgta gegaceettt geaggeageg 1500
gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560
tgcaaaggcg gcacaacccc agtgccacgt tqtqaqttqq ataqttqtqq aaaqaqtcaa 1620
atggetetee teaagegtat teaacaaggg getgaaggat geccagaagg taccccattg 1680
tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaaa 1740
aaacgtetag geeceecgaa eeacggggad gtggttttee tttgaaaaac acgataatae 1800
c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849
  Met Asp Arg Glu Met Ala Ala Ser Cys Gly Ala Val Phe Val Gly
ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg
                                                                  1897
Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg
ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg
                                                                  1945
Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu
caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc
                                                                  1993
Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val
     50
atc etc etc acg tge geg atc cac cca gag eta atc ttt acc atc acc
                                                                  2041
Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile Thr
 65
aaa atc ttg ctc gcc ata ctc ggt cca ctc atq qtq ctc cag gct qqt
Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala Gly
                 85
ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt gca
                                                                  2137
Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg Ala
           100
                                105
```

tgc Cys	atg Met	ctg Leu 115	gtg Val	cgg Arg	aag Lys	gtt Val	gct Ala 120	eja aaa	ggt Gly	cat His	tat Tyr	gtc Val 125	caa Gln	atg <sup>°</sup> Met	gct Ala	2185
ctc Leu	atg Met 130	aag Lys	ttg Leu	gcc Ala	gca Ala	ctg Leu 135	aca Thr	ggt Gly	acg Thr	tac Tyr	gtt Val 140	tat Tyr	gac Asp	cat His	ctc Leu	2233
				gac Asp												2281
				gtc Val 165												2329
tgg Trp	gjå aaa	gca Ala	gac Asp 180	acc Thr	gcg Ala	gcg Ala	tgt Cys	999 Gly 185	gac Asp	atc Ile	atc Ile	ttg Leu	ggc Gly 190	ctg Leu	ccc Pro	2377
gtc Val	tcc Ser	gcc Ala 195	cgc Arg	agg Arg	gly aaa	agg Arg	gag Glu 200	ata Ile	cat His	ctg Leu	gga Gly	ccg Pro 205	gca Ala	gac Asp	agc Ser	2425
ctt Leu	gaa Glu 210	gly ggg	cag Gln	999 999	tgg Trp	cga Arg 215	ctc Leu	ctc Leu	gcg Ala	cct Pro	att Ile 220	acg Thr	gcc Ala	tac Tyr	tcc Ser	2473
caa Gln 225	cag Gln	acg Thr	cga Arg	ggc Gly	cta Leu 230	ctt Leu	ggc	tgc Cys	atc Ile	atc Ile 235	acc Thr	agc Ser	ctc Leu	aca Thr	ggc Gly 240	2521
cgg Arg	gac Asp	agg Arg	aac Asn	cag Gln 245	gtc Val	gag Glu	gjå aaa	gag Glu	gtc Val 250	caa Gln	gtg Val	gtc Val	tcc Ser	acc Thr 255	gca Ala	2569
				ctg Leu												2617
				ggc Gly												2665
				acc Thr												2713
				cgt Arg												2761
ctt Leu	tac Tyr	ttg Leu	gtc Val	acg Thr 325	aag Lys	cat His	gcc Ala	gat Asp	gtc Val 330	att Ile	ccg Pro	gtg Val	cgc Arg	cgg Arg 335	cgg Arg	2809

					agc Ser											2857
_					ggt Gly		_		_		_			_		2905
					gcc Ala											2953
gac Asp 385	ttt Phe	gta Val	ccc Pro	gtc Val	gag Glu 390	tct Ser	atg Met	gaa Glu	acc Thr	act Thr 395	atg Met	cgg Arg	tcc Ser	ccg Pro	gtc Val 400	3001
		_		_	tcc Ser		-	_	_	_	_			_		3049
					cct Pro											3097
					caa Gln											3145
					ggt Gly											3193
Ile 465	Asp	Pro	Asn	Ile	aga Arg 470	Thr	Gly	Val	Arg	Thr 475	Ile	Thr	Thr	Gly	Ala 480	3241
Pro	Ile	Thr	Tyr	Ser 485	acc Thr	Tyr	Gly	Lys	Phe 490	Leu	Ala	Asp	Gly	Gly 495	Cys	3289
Ser	Gly	Gly	Ala 500	Tyr	gac Asp	Ile	Ile	Ile 505	Cys	Asp	Glu	Cys	His 510	Ser	Thr	3337
					ctg Leu											3385
					ctc Leu											3433
					cat His 550											3481

atc ccc ttt Ile Pro Phe 565				Ile
agg cac ctc Arg His Leu 580				
gcg aag ctg Ala Lys Leu				
 ctt gat gta Leu Asp Val	-	_		
 acg gac gct Thr Asp Ala 630				
gac tgc aat Asp Cys Asn 645				e Ser
acc ttc acc Thr Phe Thr 660		Thr Thr Val		
tcg cag cgg Ser Gln Arg				
 ttt gtg act Phe Val Thr				
ctg tgc gag Leu Cys Glu 710				
 gcc gag acc Ala Glu Thr		Leu Arg Ala		Thr ,
725 ccc gtc tgc Pro Val Cys 740		Leu Glu Phe	tgg gag ggc	gtc 4057
ctc acc cac Leu Thr His				
gac aac ttc Asp Asn Phe				
gct cag gct Ala Gln Ala 790				

												acg Thr				4249
		_		-	_				-			aca Thr				4297
												gag Glu 845				4345
_		~~		_	_			-		_	-	ctg Leu	_			4393
												atc Ile				4441
		_	_				~			_		tac Tyr				4489
_		_	_		_	_						atc Ile	_	_	gga Gly	4537
	_		_	_				_				999 Gly 925	_	_		4585
							_	-	_			gtg Val	-		_	4633
				_	_				_		-	tgg Trp				4681
_						_	~~	-			_	cct Pro				4729
												acc Thr				47,77
			His			_		Asn		_		gga Gly 1005	Trp		_	4825
		ctt Leu					Āla					gta Val				4873

11e Ala Giy	gcg gct Ala Ala		_		Gly Lys			L
gat att tto Asp Ile Lev		Tyr Gly					Ala	Э
ttt aag gto Phe Lys Val			_	Ser Thr				7
cta ctc cct Leu Leu Pro 107	Ala Ile					Gly Val		5
tgc gca gcg Cys Ala Ala 1090			His Val					3
cag tgg atg Gln Trp Met 1105					Arg Gly			L
tcc ccc acc Ser Pro Thi		Val Pro					Thr	Э
cag atc cto Gln Ile Leu	_			Gln Leu	-			7
ana taa nto	. 220 020							_
Gln Trp Ile	Asn Glu		_	_	tcc ggc Ser Gly 116	Ser Trp		5
Gln Trp Ile	Asn Glu 5 tgg gat	Asp Cys	Ser Thr 1160 tgc acg Cys Thr	Pro Cys	Ser Gly 116! act gat	Ser Trp ttc aag	Leu gcc 5353	
Gln Trp Ile 115 aga gat gtt Arg Asp Val	Asn Glu  tgg gat Trp Asp	Asp Cys  tgg ata Trp Ile 117! ctc ctg	ser Thr 1160  tgc acg Cys Thr  ccg cga	Pro Cys gtg ttg Val Leu ttg ccg	ser Gly 1169 act gat Thr Asp 1180 gga gtc Gly Val	ser Trp ttc aag Phe Lys ccc ttc	gcc 5353 Ala ttc 5403	3
Gln Trp Ile 115 aga gat gtt Arg Asp Val 1170 tgg ctc cag Trp Leu Glr	Asn Glu tgg gat Trp Asp tcc aag Ser Lys	Asp Cys  tgg ata Trp Ile     1179  ctc ctg Leu Leu 1190  tac aag Tyr Lys	tgc acg Cys Thr ccg cga Pro Arg	gtg ttg Val Leu ttg ccg Leu Pro 1199	ser Gly 1169 act gat Thr Asp 1180 gga gtc Gly Val 5	Ser Trp  ttc aag Phe Lys  ccc ttc Pro Phe  ggc atc	gcc 5353 Ala 5403 Phe 1200 atg 5443 Met	3
aga gat gtt Arg Asp Val 1170  tgg ctc cag Trp Leu Glr 1185  tca tgt cas	Asn Glu  tgg gat  trp Asp  tcc aag  Ser Lys  cgt ggg Arg Gly  tgc cca	Asp Cys  tgg ata Trp Ile 117! ctc ctg Leu Leu 1190  tac aag Tyr Lys  tgt gga	tgc acg Cys Thr Ccg cga Pro Arg gga gtc Gly Val	gtg ttg Val Leu ttg ccg Leu Pro 1199 tgg cgg Trp Arg 1210 atc acc Ile Thr	ser Gly 1169 act gat Thr Asp 1180 gga gtc Gly Val 5 ggc gac Gly Asp	Ser Trp  ttc aag Phe Lys  ccc ttc Pro Phe  ggc atc Gly Ile 1219 gtg aaa	gcc 5353 Ala  ttc 5403 Phe 1200 atg 5449 Met 5	3 1

gga aca ttc ccc att Gly Thr Phe Pro Ile 1250			Cys Thr Pro Ser
ccg gcg cca aat tat Pro Ala Pro Asn Tyr 1265			
tac gtg gag gtt acg Tyr Val Glu Val Thr 128	Arg Val Gly		
acc act gac aac gta Thr Thr Asp Asn Val 1300			
ttc aca gaa gtg gat Phe Thr Glu Val Asp 1315		Leu His Arg Tyr	
aaa ccc ctc cta cgg Lys Pro Leu Leu Arg 1330.			Gly Leu Asn Gln
tac ccg gtt ggg tca Tyr Pro Val Gly Ser 1345			
gtg ctc act tcc atg Val Leu Thr Ser Met 136	Leu Thr Asp		
gct aag cgt agg ctg Ala Lys Arg Arg Leu 1380			<del>-</del>
tca gct agc cag ctg Ser Ala Ser Gln Leu 1395		Ser Leu Lys Ala	
cgt cat gac tcc ccg Arg His Asp Ser Pro 1410		~ ~ ~	Asn Leu Leu Trp
cgg cag gag atg ggc Arg Gln Glu Met Gly 1425			
gta gta att ttg gac Val Val Ile Leu Asp 144	Ser Phe Glu	ccg ctc caa gcg Pro Leu Gln Ala 1450	gag gag gat gag 6169 Glu Glu Asp Glu 1455
agg gaa gta tcc gtt Arg Glu Val Ser Val 1460	~ w w w		

cct cga gcg a Pro Arg Ala M 1475						
tta gag tcc t Leu Glu Ser T 1490						
tgt cca ttg c Cys Pro Leu P 1505	-	Lys Ala Pr	-	Pro Pro P		
aag agg acg g Lys Arg Thr V						Ala
gag ctc gcc a Glu Leu Ala T 1	-	Phe Gly Se		Ser Ser A		_
agc ggc acg g Ser Gly Thr A 1555		•	~			~
gcg gga tcc g Ala Gly Ser A 1570						
gag ccg ggg g Glu Pro Gly A 1585	_	Leu Ser As		Trp Ser T	_	_
gag gag gct a Glu Glu Ala S				-		Trp
aca ggc gcc c Thr Gly Ala L		Pro Cys Al		Glu Thr L		
ato aat ooa o						
Ile Asn Ala L 1635		tct ttg ct Ser Leu Le 1640				
Ile Asn Ala L	su Ser Asn	Ser Leu Le 1640 gca agc ct	eu Arg His	His Asn L 1645 aag aag g	eu Val	Tyr ttt 6793
Ile Asn Ala L 1635 gct aca aca t Ala Thr Thr S	nu Ser Asn et ege age er Arg Ser g gte etg	Ser Leu Le 1640 gca agc ct Ala Ser Le 1655 gac gac ca Asp Asp Hi	eu Arg His cg cgg cag eu Arg Gln ac tac cgg	His Asn L 1645 aag aag g Lys Lys V 1660 gac gtg c Asp Val L	eu Val  tc acc al Thr	Tyr  ttt 6793 Phe  gag 6841

gaa gcc tgt aag ctg Glu Ala Cys Lys Leu 1700	acg ccc cca cat tcg Thr Pro Pro His Ser 1705		
tat ggg gca aag gac Tyr Gly Ala Lys Asp 1715	gtc cgg aac cta tcc Val Arg Asn Leu Ser 1720	agc aag gcc gtt a Ser Lys Ala Val A 1725	aac cac 6985 Asn His
	aag gac ttg ctg gaa Lys Asp Leu Leu Glu 1735		
gac acc acc atc atg Asp Thr Thr Ile Met 1745	gca aaa aat gag gtt Ala Lys Asn Glu Val 1750	ttc tgc gtc caa o Phe Cys Val Gln 1 1755	cca gag 7081 Pro Glu 1760
aag ggg ggc cgc aag Lys Gly Gly Arg Lys 1769	cca gct cgc ctt atc Pro Ala Arg Leu Ile 5 177	Val Phe Pro Asp 1	ttg ggg 7129 Leu Gly 1775
gtt cgt gtg tgc gag Val Arg Val Cys Glu 1780	aaa atg gcc ctt tac Lys Met Ala Leu Tyr 1785	gat gtg gtc tcc a Asp Val Val Ser 1790	acc ctc 7177 Thr Leu
	ggc tct tca tac gga Gly Ser Ser Tyr Gly 1800		
cag cgg gtc gag ttc Gln Arg Val Glu Phe · 1810	ctg gtg aat gcc tgg Leu Val Asn Ala Trp 1815	aaa gcg aag aaa i Lys Ala Lys Lys ( 1820	tgc cct 7273 Cys Pro
	gac acc cgc tgt ttt Asp Thr Arg Cys Phe 1830		
	gag gag tca atc tac Glu Glu Ser Ile Tyr 5 185	Gln Cys Cys Asp	
	gcc ata agg tcg ctc Ala Ile Arg Ser Leu 1865		
ggg ggc ccc ctg act Gly Gly Pro Leu Thr 1875	aat tot aaa ggg cag Asn Ser Lys Gly Gln 1880	aac tgc ggc tat ( Asn Cys Gly Tyr ) 1885	cgc cgg 7465 Arg Arg
tgc cgc gcg agc ggt Cys Arg Ala Ser Gly 1890	gta ctg acg acc agc Val Leu Thr Thr Ser 1895	tgc ggt aat acc o Cys Gly Asn Thr 1 1900	ctc aca 7513 Leu Thr
tgt tac ttg aag gcc 'Cys Tyr Leu Lys Ala 1905	gct gcg gcc tgt cga Ala Ala Ala Cys Arg 1910	gct gcg aag ctc o Ala Ala Lys Leu o 1915	cag gac 7561 Gln Asp 1920

tgc acg a	Met Leu					val.					Ser	7609
gcg ggg a		Glu Asp			Leu					Glu		7657
atg act a Met Thr A	-	_		Gly	_	-			Pro	-		7705
gac ttg g Asp Leu ( 1970								Ser				7753
gat gca t Asp Ala 8 1985			Val Tyr				Arg	_				7801
ccc ctt c Pro Leu A	Ala Arg I					Arg					Asn	7849
tcc tgg o Ser Trp l					Ala					Āla		7897
atg atc o				Ser				_	Gln	_		7945
ctt gaa a Leu Glu I 2050	_	_						Cys				7993
gag cca o Glu Pro I 2065	_		Gln Ile			_	Leu				_	8041
gca ttt t Ala Phe S	Ber Leu I					Glu					Ala	8089
tca tgc o Ser Cys I					Pro					Arg		8137
cgg gcc a Arg Ala A				Leu					Gly			8185
gcc act t Ala Thr 0 2130								Arg				8233

	Leu					Ala					Asp	tta Leu				8281
ttc Phe	gtt Val	gct Ala	ggt Gly	tac Tyr 2165	Ser	gly ggg	gga Gly	gac Asp	ata Ile 2170	Tyr	cac His	agc Ser	ctg Leu	tct Ser 2175	Arg	8329
				$\mathtt{Trp}$					Leu			ctt Leu		Val		8377
-			Tyr			ccc Pro		Arg		acgg	ggag	jct a	laaca	actéo	a	8427
tttt	tttt ctta	tt t	ctte	tcct	t tt	tttt	tect tgtg	ctt aaa	tttt	tcc	tttt	cttt	cc t	ttgg	ttttt gtggct cagaga	8547